

Pending Nucleic Acid and Pending Amino Acid database searches generate two sets of results each. The Pending databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions **.rapm** and **.rapn**

Because they contain data that is confidential, the results of Pending database searches should not be left in the case .

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:15:59 ; Search time 34.1497 Seconds
(without alignments)
2367.017 Million cell updates/sec

Title: US-10-658-989A-1

Perfect score: 1184

Sequence: 1 GPGEPTGLPQPPGERGG.....GEQVPGDLGAPGSPAGG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_l6Dec04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003as.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1184	100.0	209	ADM48390	Adm48390 Human rec
2	1178	99.5	617	ADM48391	Adm48391 Human rec
3	1178	99.5	617	ADQ26217	Adq26217 Human gel
4	1178	99.5	821	ADM48392	Adm48392 Human rec
5	1149	97.0	501	AAB68057	Aab68057 Amino aci
6	1149	97.0	501	AAE02703	Aae02703 Human alp
7	1149	97.0	501	ADB84290	Adb84290 Recombina
8	1149	97.0	1057	AAH84541	Aah84541 Amino aci
9	1149	97.0	1057	AAH84544	Aah84544 A human c
10	1149	97.0	1058	AAH84403	Aah84403 Amino aci
11	1149	97.0	1107	AAH89472	Aah89472 Collagen/
12	1149	97.0	1107	AAH84540	Aah84540 Amino aci
13	1149	97.0	1161	AAH87050	Aah87050 Human pan
14	1149	97.0	1169	AAH89469	Aah89469 Collagen/
15	1149	97.0	1169	AAH84537	Aah84537 Amino aci
16	1149	97.0	1171	AAH89470	Aah89470 Collagen/
17	1149	97.0	1171	AAH84538	Aah84538 A chimeri
18	1149	97.0	1211	AAH87057	Aah87057 Human pan
19	1149	97.0	1226	AAH87062	Aah87062 Human pan
20	1149	97.0	1341	AAH71701	Aah71701 Collagen
21	1149	97.0	1341	AAH96122	Aah96122 Collagen
22	1149	97.0	1341	AAE16475	Aae16475 Human col
23	1149	97.0	1341	ABH80733	Abh80733 Collagen
24	1149	97.0	1341	ABH09625	Abh09625 Amino aci
25	1149	97.0	1341	ADF13075	Adf13075 Human col

26	1149	97.0	1388	2	AAR89471	Aar89471 Collagen/
27	1149	97.0	1411	3	AAH56800	Aah56800 Human pre
28	1149	97.0	1461	5	ABG33947	Abg33947 Human pol
29	1149	97.0	1464	2	AAW68485	Aaw68485 Human rec
30	1149	97.0	1464	4	AAH82454	Aah82454 Human pro
31	1149	97.0	1464	4	AAU14136	Aau14136 Human nov
32	1149	97.0	1464	5	ABB90764	Abb90764 Human tum
33	1149	97.0	1464	5	ABP68610	Abp68610 Human pan
34	1149	97.0	1464	6	ABU54471	Abu54471 Human tum
35	1149	97.0	1464	6	ABR47417	AbR47417 Breast ca
36	1149	97.0	1464	6	ABR32064	AbR32064 Human cer
37	1149	97.0	1464	7	ADD14142	Add14142 Human src
38	1149	97.0	1464	7	ADD45059	Add45059 Human pro
39	1149	97.0	1464	7	ADD45055	Add45055 Human pro
40	1149	97.0	1464	7	ADD45051	Add45051 Human pro
41	1149	97.0	1464	7	ADP65246	Adp65246 Human alp
42	1149	97.0	1464	7	ADQ19470	Adq19470 Human sof
43	1149	97.0	1464	8	ADQ29653	Adq29653 Human col
44	1149	97.0	1464	8	ADR16800	Adr16800 Human col
45	1149	97.0	1464	8	ADR16800	Adr16800 Human col

ALIGNMENTS

RESULT 1

ADM48390

ID ADM48390 standard; protein; 209 AA.

XX AC ADM48390;

XX DT 03-JUN-2004 (first entry)

XX DE Human recombinant gelatin-like polypeptide Hu-1.

XX KW Plasma substitute; Gelatin-like protein; plasma expander; human.

XX OS Homo sapiens.

XX PN EP1398324-A1.

XX PD 17-MAR-2004.

XX PF 11-SEP-2002; 2002EP-00078745.

XX PR 11-SEP-2002; 2002EP-00078745.

XX PA (FUJF) FUJI PHOTO FILM BV.

XX PI Bouwstra JB, Toda Y;

XX WPI; 2004-229415/22.

Composition useful as substitute for plasma, comprises solution of saline and recombinant gelatin-like protein having colloid osmotic function.

Example 1; SEQ ID NO 1; 31pp; English.

The present sequence is the protein sequence of human recombinant gelatin-like protein Hu-1. This has a molecular weight of 18.4 kDa and an isoelectric point of 5.35. A claimed composition suitable as a substitute for plasma comprises a solution of saline and a protein having a colloid osmotic function. The protein is a recombinant gelatin-like protein with a molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric point of less than 8. It is especially Hu-1 or Hu-deam ADM48393. The recombinant gelatin-like protein, or a dimer, trimer or tetramer of the protein, is useful as a plasma expander that has a lower clearance rate from blood circulation, provides better and predictable regulation of clearance rate and which is less susceptible to proteolytic degradation than presently used gelatin derivatives. Recombinant gelatin-like proteins that are in essence free of hydroxyproline do not give rise to an immunological reaction with blood samples containing IgE antibodies. The gelatin-like proteins can be covalently attached to

PD 17-MAY-2001.
 XX 10-NOV-2000; 2000WO-US030843.
 XX 12-NOV-1999; 99US-0165114P.
 PR 15-MAY-2000; 2000US-0204437P.
 XX (FIBR-) FIBROGEN INC.
 XX Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
 XX WPI; 2001-308784/32.
 XX Vaccine formulations (I) comprising recombinant human gelatin, useful for
 PT vaccinating against e.g. mumps, measles, rubella, tetanus, rabies and
 PT cholera, the gelatin is non-immunogenic and confers stability at ambient
 PT temperatures.
 XX Claim 11; Page 114-116; 130pp; English.
 XX The present sequence represents a human recombinant gelatin polypeptide.
 CC The recombinant gelatin polypeptide is used to produce vaccine
 CC formulations of the invention. The recombinant human gelatin is non-
 CC immunogenic (therefore reducing anaphylactic reactions) and confers
 CC stability at ambient temperatures. The vaccine formulation comprises a
 CC vaccine formulated for the prevention of a disease selected from vaccinia
 CC virus (small pox), polio virus (Salk and Sabin), mumps, measles, rubella,
 CC diphtheria, tetanus, Varicella-Zoster (chicken pox/shingles), pertussis
 CC (whooping cough), Bacille Calmette-Guérin (BCG, tuberculosis),
 CC haemophilus influenzae meningitis, rabies, cholera, Japanese
 CC encephalitis virus, salmonella typhi, shigella, hepatitis A and B,
 CC adenovirus, yellow fever, foot and mouth disease, herpes simplex virus,
 CC respiratory syncytial virus, rotavirus, Dengue, West Nile virus, turkey
 CC herpes virus (Marek's disease), influenza and/or anthrax
 XX Sequence 501 AA;
 SQ
 Query Match 97.0%; Score 1149; DB 4; Length 501;
 Best Local Similarity 93.6%; Pred. No. 3.7e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPGTGLPGPPGGRGSGRFGGADGVAGPKPAGERSGPGA 48
 DB 280 GPPGAGEGKRGARGEPGTGLPGPPGGRGSGRFGGADGVAGPKPAGERSGPGA 339
 QY 49 GPKSGPGEAGRPEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
 DB 340 GPKSGPGEAGRPEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 399
 QY 109 GVMGFPKGAAGEPKAGRGVPGPPGAVGAGKDGAGAGQPPGPPGAGERGEGQGA 168
 DB 400 GVMGFPKGAAGEPKAGRGVPGPPGAVGAGKDGAGAGQPPGPPGAGERGEGQGA 459
 QY 169 GSPGFQGLPGPAGPGEAGKPGEGVPGDLGAPGSGPAG 208
 DB 460 GSPGFQGLPGPAGPGEAGKPGEGVPGDLGAPGSGARG 499
 RESULT 6
 AA02703
 ID AA02703 standard; protein; 501 AA.
 XX AA02703;
 XX 06-AUG-2001 (first entry)
 XX Human alpha (I) type I collagen helical domain (residues 179-679).
 KW Human; recombinant gelatin; binding agent; stabilising agent; emulsifier;
 KW encapsulant; film-forming agent; moisturising agent; thickening agent;
 KW gelling agent; colloidal agent; adhesive agent; gel capsule; photography;
 KW plasma expander; colloidal volume replacement material; graft coating;
 KW medical sponge; medical plug; micro-carrier; edible composition;

KW protein supplement; fat substitute; nutritional supplement; cell culture;
 KW edible coating; cosmetic; vaccine; therapy; arthritis; achrosis;
 KW cartilage degeneration; joint flexibility; food industry; beverage;
 KW alpha (I) type I collagen.
 OS Homo sapiens.
 XX WO200134646-A2.
 XX 17-MAY-2001.
 PD 10-NOV-2000; 2000WO-US030791.
 XX 12-NOV-1999; 99US-0165114P.
 PR 15-MAY-2000; 2000US-0204437P.
 XX (FIBR-) FIBROGEN INC.
 XX Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
 XX WPI; 2001-329072/34.
 XX Gelatin useful for pharmaceuticals, cosmetics and edible foods, is
 PT prepared recombinantly.
 XX Claim 21; Page 121-123; 137pp; English.
 XX The patent discloses recombinant human gelatin which is useful in various
 CC compositions including binding agents, encapsulants, stabilising agents,
 CC film-forming agents, moisturising agents, emulsifiers, thickening agents,
 CC gelling agents, colloidal agents, adhesive agents, pharmaceutical
 CC compositions, hard gel capsules, soft gel capsules, plasma expander,
 CC colloidal volume replacement materials, graft coatings, medical sponges,
 CC medical plugs, pharmaceutical stabilisers, micro-carriers, edible
 CC compositions, protein supplements, fat substitutes, nutritional
 CC supplements, edible coatings, photographic compositions, cosmetic
 CC compositions, industrial composition, cell culture compositions and
 CC compositions for use in the laboratory. Pharmaceutical compositions
 CC comprising recombinant gelatin are used as vaccines. They are also used
 CC to treat various joint conditions such as arthritis, achrosis and other
 CC conditions related to the degeneration of cartilage and joint
 CC flexibility. Recombinant gelatin is also used in food and beverage
 CC industries. The present sequence is human alpha (I) type I collagen
 CC helical domain (residues 179-679). This sequence is a recombinant gelatin
 XX Sequence 501 AA;
 SQ
 Query Match 97.0%; Score 1149; DB 4; Length 501;
 Best Local Similarity 93.6%; Pred. No. 3.7e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPGTGLPGPPGGRGSGRFGGADGVAGPKPAGERSGPGA 48
 DB 280 GPPGAGEGKRGARGEPGTGLPGPPGGRGSGRFGGADGVAGPKPAGERSGPGA 339
 QY 49 GPKSGPGEAGRPEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
 DB 340 GPKSGPGEAGRPEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 399
 QY 109 GVMGFPKGAAGEPKAGRGVPGPPGAVGAGKDGAGAGQPPGPPGAGERGEGQGA 168
 DB 400 GVMGFPKGAAGEPKAGRGVPGPPGAVGAGKDGAGAGQPPGPPGAGERGEGQGA 459
 QY 169 GSPGFQGLPGPAGPGEAGKPGEGVPGDLGAPGSGPAG 208
 DB 460 GSPGFQGLPGPAGPGEAGKPGEGVPGDLGAPGSGARG 499
 RESULT 7
 AD84290
 ID AD84290 standard; protein; 501 AA.
 XX AD84290;
 AC AD84290;

XX 04-DEC-2003 (first entry)
 XX Recombinant gelatin #3.
 XX virucide; antitussive; tuberculostatic; antibacterial; antiinflammatory;
 KW hepatotropic; vaccine; recombinant gelatin; vaccinia virus; small pox;
 KW polio virus; Salk and Sabin; mumps; measles; rubella; diphtheria;
 KW tetanus; Varicella-Zoster; chicken pox; shingles; pertussis;
 KW whooping cough; Bacille Calmette-Guerin; BCG; tuberculosis;
 KW Haemophilus influenzae; meningitis; rabies; cholera;
 KW Japanese encephalitis virus; Salmonella typhi; Shigella; hepatitis A;
 KW hepatitis B; adenovirus; yellow fever; foot and mouth disease;
 KW herpes simplex virus; respiratory syncytial virus; rotavirus; dengue;
 KW West Nile virus; turkey herpes virus; Marek's disease; influenza;
 KW anthrax; collagen type 1; alpha1(I).
 XX Synthetic.
 OS Homo sapiens.
 XX US2003064074-A1.
 XX 03-APR-2003.
 XX 30-AUG-2002; 2002US-00232175.
 XX 12-NOV-1999; 99US-0165114P.
 PR 15-MAY-2000; 2000US-0204437P.
 PR 10-NOV-2000; 2000US-00710249.
 XX (CHAN/) CHANG R C.
 PA (KIVI/) KIVIRIKKO K I.
 PA (NEFF/) NEFF T B.
 PA (OLSE/) OLSEN D R.
 PA (POLA/) POLAREK J W.
 XX Chang RC, Kivirikko KI, Neff TB, Olsen DR, Polarek JW;
 XX WPI; 2003-540775/51.
 XX New vaccine composition comprising a recombinant gelatin and an antigenic
 PT agent, useful for preventing e.g. polio virus, mumps, measles, rubella,
 PT diphtheria, tetanus, chicken pox/shingles, pertussis, cholera, rotavirus
 PT or dengue.
 XX Claim 16; Page 36-38; 63pp; English.
 PS The invention describes a vaccine composition comprising a recombinant
 CC gelatin, and an antigenic agent. The vaccine can be delivered by
 CC injection, through nasal, oral, transdermal or mucosal routes, or through
 CC deep lung delivery. Administration may also be oral, rectal,
 CC transcutaneous, intramedullary, intrathecal, intraventricular, intravenous,
 CC intraperitoneal, intranasal or intraocular injection). The vaccine
 CC composition is formulated for the prevention of vaccinia virus (small
 CC pox), polio virus (Salk and Sabin), mumps, measles, rubella, diphtheria,
 CC tetanus, Varicella-Zoster (chicken pox/shingles), pertussis (whooping
 CC cough), Bacille Calmette-Guerin (BCG, tuberculosis), Haemophilus
 CC influenzae meningitis, rabies, cholera, Japanese encephalitis virus,
 CC Salmonella typhi, Shigella hepatitis A, hepatitis B, adenovirus, yellow
 CC fever, foot and mouth disease, herpes simplex virus, respiratory
 CC syncytial virus, rotavirus, dengue, West Nile virus, Turkey herpes virus
 CC (Marek's disease), influenza, and anthrax. This is the amino acid
 CC sequence of a recombinant gelatin used in the creation of a vaccine
 CC composition of the invention.
 XX Sequence 501 AA;
 SQ Query Match 97.0%; Score 1149; DB 7; Length 501;
 Best Local Similarity 93.6%; Pred. No. 3.7e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPGPTGLPGPGERGPGSRGFPAGDVGAGPKGAGERGSGPGA 48

Db 280 GPGPAGAGEBKGARGEPGPTGLPGPGERGPGSRGFPAGDVGAGPKGAGERGSGPGA 339
 Qy 49 GPKGSPGEGACRPGEGAGLPGAKGLTSGSPGPGPKTGTGPPAGQDGRGPPGPPGARGQA 108
 Db 340 GPKGSPGEGACRPGEGAGLPGAKGLTSGSPGPGPKTGTGPPAGQDGRGPPGPPGARGQA 399
 Qy 109 GVMGFPKGAAGEPGKAGRGVGPBGAVGPAGKDGEGAGAQGPPGAGPAGERGEGOPA 168
 Db 400 GVMGFPKGAAGEPGKAGRGVGPBGAVGPAGKDGEGAGAQGPPGAGPAGERGEGOPA 459
 Qy 169 GSPGFQGLPGPAGPPGPGGAGKPGEGQGVPGDLGAPGSPGAG 208
 Db 460 GSPGFQGLPGPAGPPGPGGAGKPGEGQGVPGDLGAPGSPGAG 499
 RESULT 8
 AAY84541
 ID AAY84541 standard; protein; 1057 AA.
 XX
 AC AAY84541;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE Amino acid sequence of a human collagen 1 (alpha1) protein.
 KW Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibrinogen; fibronectin; post translational hydroxylation.
 XX
 OS Homo sapiens.
 XX
 PN EP992586-A2.
 XX
 PD 12-APR-2000.
 XX
 PF 07-OCT-1999; 99EP-00119184.
 XX
 PR 09-OCT-1998; 98US-00169768.
 XX
 PA (USSU) US SURGICAL CORP.
 XX
 PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
 XX
 DR WPI; 2000-259138/23.
 DR N-PSDB; AAA12502.
 XX
 PT Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 PS
 XX Disclosure; Fig 27A-E; 260pp; English.
 CC The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a human collagen 1 (alpha1) protein, which
 CC may be produced using the method of the invention

XX SQ Sequence 1057 AA;
Query Match 97.0%; Score 1149; DB 3; Length 1057;
Best Local Similarity 93.6%; Pred. No. 6.8e-67;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GSPGPTGLPGPPGGRGGPSRFPAGDVGAGPKGPPAGRGSPGPA 48
DB 297 GPPGAGEGKRGARGEGPTGLPGPPGGRGGPSRFPAGDVGAGPKGPPAGRGSPGPA 356
QY 49 GPKSGPGEAGRPAGLPGAKLTGSPGSPGDPKTPGPPAGDGRPPGPPGARGQA 108
DB 357 GPKSGPGEAGRPAGLPGAKLTGSPGSPGDPKTPGPPAGDGRPPGPPGARGQA 416
QY 109 GVMGFPKGAAAGPAGRGVPPGCAVGPAGKDGAGAGGPPGPPAGRGSGGPA 168
DB 417 GVMGFPKGAAAGPAGRGVPPGCAVGPAGKDGAGAGGPPGPPAGRGSGGPA 476
QY 169 GSPGFQGLPGPPGAGKRGVPPGCDLGA GPSPGAG 208
DB 477 GSPGFQGLPGPPGAGKRGVPPGCDLGA GPSPGAG 516
RESULT 9
AAY84544
ID AAY84544 standard; protein; 1057 AA.
AC AAY84544;
XX
DT 25-JUL-2000 (first entry)
XX
DE A human collagen 1 (alpha) protein helical region.
XX
KW Extracellular matrix protein; self aggregation; hydroxylated proline;
trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX
OS Homo sapiens.
XX
PN EP992586-A2.
XX
PD 12-APR-2000.
XX
PF 07-OCT-1999; 99EP-00119184.
XX
PR 09-OCT-1998; 98US-00169768.
XX
PA (USSU) US SURGICAL CORP.
XX
PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX
DR WPI; 2000-259138/23.
XX
DR N-PSDB; AAA12503.
XX
PT Production of extracellular matrix proteins containing 4-trans-
hydroxyproline results in native self aggregating proteins, useful on
PT medical implants.
XX
PS Example 10; Fig 39A-E; 260pp; English.
XX
CC The specification describes a method for producing an extracellular
matrix protein or its fragment. The extracellular matrix protein is
CC capable of self aggregating in a cell which does not ordinarily
CC hydroxylated prolines. The method comprises optimising a nucleic acid
CC sequence for expression in the cell by substitution of codons preferred
CC by that cell for naturally occurring codons not preferred by the cell;
CC incorporating the nucleic acid sequence into the cell; and contacting the
CC cell with a hypertonic growth medium containing at least one amino acid,
CC selected from the group consisting of trans-4-hydroxyproline and 3-
CC hydroxyproline to allow at least one of the amino acids to be assimilated
CC into the cell and incorporated into the extracellular matrix protein. The
CC method may be used to make host cells assimilate and incorporate trans-4-

CC hydroxyproline into proteins. This is especially useful in the
CC recombinant production of proteins such as collagen, fibrinogen and
CC fibronectin whose ability to self aggregate and produce functional
CC proteins depends on the post translational hydroxylation of proline. The
CC method is also useful in studying the structure and function of
CC polypeptides which do not normally contain trans-4-hydroxyproline. The
CC present sequence represents human collagen 1 (alpha) helical region,
CC which may be produced using the method of the invention
XX
SQ Sequence 1057 AA;
Query Match 97.0%; Score 1149; DB 3; Length 1057;
Best Local Similarity 93.6%; Pred. No. 6.8e-67;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GSPGPTGLPGPPGGRGGPSRFPAGDVGAGPKGPPAGRGSPGPA 48
DB 297 GPPGAGEGKRGARGEGPTGLPGPPGGRGGPSRFPAGDVGAGPKGPPAGRGSPGPA 356
QY 49 GPKSGPGEAGRPAGLPGAKLTGSPGSPGDPKTPGPPAGDGRPPGPPGARGQA 108
DB 357 GPKSGPGEAGRPAGLPGAKLTGSPGSPGDPKTPGPPAGDGRPPGPPGARGQA 416
QY 109 GVMGFPKGAAAGPAGRGVPPGCAVGPAGKDGAGAGGPPGPPAGRGSGGPA 168
DB 417 GVMGFPKGAAAGPAGRGVPPGCAVGPAGKDGAGAGGPPGPPAGRGSGGPA 476
QY 169 GSPGFQGLPGPPGAGKRGVPPGCDLGA GPSPGAG 208
DB 477 GSPGFQGLPGPPGAGKRGVPPGCDLGA GPSPGAG 516
RESULT 10
AAY84403
ID AAY84403 standard; protein; 1058 AA.
XX
AC AAY84403;
XX
DT 12-JUL-2000 (first entry)
XX
DE Amino acid sequence of human type 1 (alpha) collagen polypeptide.
XX
KW Alpha collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;
collagen; mussel adhesive protein; bioadhesive.
XX
OS Homo sapiens.
XX
PN WO200014201-A1.
XX
PD 16-MAR-2000.
XX
PF 07-SEP-1999; 99WO-US020462.
XX
PR 09-SEP-1998; 98US-0099652P.
XX
PA (USSU) US SURGICAL CORP.
PA (PAOL/) PAOLELLA D N.
PA (GRUS/) GRUSKIN E A.
PA (BUEC/) BUECHTER D D.
XX
PI Paolella DN, Gruskin EA, Buechter DD;
XX
DR WPI; 2000-271051/23.
XX
DR N-PSDB; AAZ99843.
XX
PT Incorporating non-natural amino acid into polypeptide, useful e.g. for
production of bioadhesives, by epoxidation or substitution of
PT dehydroproline residues.
XX
PS Disclosure; Fig 6; 66pp; English.
XX
CC The present sequence represents a human type 1 (alpha) collagen protein.
Peptides derived from the protein were used to demonstrate incorporation

of 3,4-dehydro-L-proline into the peptide, using the method of the invention. The specification describes a method for the incorporation of non-natural amino acid into a polypeptide. The method comprises reacting at least one 3,4-dehydroproline residue in the polypeptide with an epoxidation reagent from a polypeptide containing at least one 3,4-epoxyproline residue. The method is used for studying the effects of non-natural amino acids on structure and function of polypeptides. The method is also useful for commercial production of collagen or mussel adhesive proteins (which are useful as bioadhesives), and for incorporating a wide variety of groups, including therapeutic ligands and biological probes, into polypeptides

XX Sequence 1058 AA;
 CC Query Match 97.0%; Score 1149; DB 3; Length 1058;
 CC Best Local Similarity 93.6%; Pred. No. 6.8e-67;
 CC Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPGPTGLPGERGSGRPGPGADGVAGPKGAPGERSGCPGA 48
 DB 298 GPPGAGEGKRGARPEGPTGLPGERGSGRPGPGADGVAGPKGAPGERSGCPGA 357
 QY 49 GPKGSGEAGRPGAGLPGAKGLTSGSGSPGPDGKTGTPPGAGQDGRPPGPPGARGQA 108
 DB 358 GPKGSGEAGRPGAGLPGAKGLTSGSGSPGPDGKTGTPPGAGQDGRPPGPPGARGQA 417
 QY 109 GVMGFPKGAAGSPGKAGRGVPGPGAVGAPGAKGDEAGAQGPAGPAGERGEOGPA 168
 DB 418 GVMGFPKGAAGSPGKAGRGVPGPGAVGAPGAKGDEAGAQGPAGPAGERGEOGPA 477
 QY 169 GSPGFQGLPGAPGPGGKAGEGQVPGDLGAPGSPGAG 208
 DB 478 GSPGFQGLPGAPGPGGKAGEGQVPGDLGAPGSPGARG 517

RESULT 11
 AAR89472
 ID AAR89472 standard; protein; 1107 AA.
 XX AC AAR89472;
 XX DT 01-OCT-1996 (first entry)
 XX DE Collagen/decorin(aa46-93) fusion protein.
 XX KW Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;
 XX KW bone formation; tissue repair; fusion protein.
 XX OS Synthetic.
 XX FH Key Location/Qualifiers
 XX FT Domain 1..1057
 FT /label= Collagen-IA
 FT /note= "collagen IA alpha-helical domain"
 FT Misc-difference 887
 FT /note= "unidentified amino acid"
 FT Misc-difference 890
 FT /note= "unidentified amino acid"
 FT Peptide 1058..1059
 FT /label= Linker_peptide
 FT Domain 1060..1107
 FT /label= Decorin
 FT /note= "amino acids P46 to G93 of mature decorin"
 XX CA2151547-A.
 XX PD 11-DEC-1995.
 XX PF 12-JUN-1995; 95CA-02151547.
 XX PR 10-JUN-1994; 94US-00259263.
 XX PA (USSU) US SURGICAL CORP.

XX Gruskin EA, Espino P;
 XX WI; 1996-140144/15.
 DR N-PSDB; AAT16518.
 XX Chimaeric DNA encoding protein contg. extracellular matrix protein domain
 PT - and cellular regulatory factor domain, partic. useful as osteogenic
 FT agents, also related vectors, transformed cells and chimaeric proteins.
 XX Disclosure; Fig 8; 59pp; English.

XX A fusion protein (AAR89472) comprises the alpha-helical region of human
 CC collagen I(a) linked to amino acids 46-93 of human mature dermatan
 CC sulphate proteoglycan (decorin). It can be expressed in Escherichia coli
 CC transformants carrying a vector incorporating a chimeric gene (AAT16518)
 CC coding for the fusion. The decorin binds to type I collagen and thus
 CC affects fibril formation. It inhibits the cell attachment-promoting
 CC activity of collagen and fibrinogen by binding to such molecules near
 CC their cell binding sites. The collagen moiety provides an integral
 CC substratum or scaffolding for the decorin. The fusion protein acts to
 CC reduce scarring of healing tissue
 XX Sequence 1107 AA;

Query Match 97.0%; Score 1149; DB 2; Length 1107;
 Best Local Similarity 93.6%; Pred. No. 7.1e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPGPTGLPGERGSGRPGPGADGVAGPKGAPGERSGCPGA 48
 DB 297 GPPGAGEGKRGARPEGPTGLPGERGSGRPGPGADGVAGPKGAPGERSGCPGA 356
 QY 49 GPKGSGEAGRPGAGLPGAKGLTSGSGSPGPDGKTGTPPGAGQDGRPPGPPGARGQA 108
 DB 357 GPKGSGEAGRPGAGLPGAKGLTSGSGSPGPDGKTGTPPGAGQDGRPPGPPGARGQA 416
 QY 109 GVMGFPKGAAGSPGKAGRGVPGPGAVGAPGAKGDEAGAQGPAGPAGERGEOGPA 168
 DB 417 GVMGFPKGAAGSPGKAGRGVPGPGAVGAPGAKGDEAGAQGPAGPAGERGEOGPA 476
 QY 169 GSPGFQGLPGAPGPGGKAGEGQVPGDLGAPGSPGAG 208
 DB 477 GSPGFQGLPGAPGPGGKAGEGQVPGDLGAPGSPGARG 516

RESULT 12
 AAY84540
 ID AAY84540 standard; protein; 1107 AA.
 XX AC AAY84540;
 XX DT 25-JUL-2000 (first entry)
 XX DE Amino acid sequence of a chimeric collagen 1 (alpha1)/decorin protein.
 XX KW Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibrinogen; fibronectin; post translational hydroxylation;
 KW decorin; chimera.
 OS Homo sapiens.
 OS Unidentified.
 OS Chimeric.
 XX Key Location/Qualifiers
 XX FT Misc-difference 858
 FT /note= "Gly encoded by GCT"
 XX EP992586-A2.
 XX PD 12-APR-2000.
 XX

PF 07-OCT-1999; 99EP-00119184.
 XX 09-OCT-1998; 98US-00169768.
 XX (USSU) US SURGICAL CORP.
 XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
 XX WPI; 2000-259138/23.
 DR N-PSDB; AAA12500.
 XX
 PT Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX
 PS Claim 24; Fig 18; 260pp; English.
 CC The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a chimeric collagen 1 (alpha1)/decorin
 CC protein, which may be produced using the method of the invention
 XX
 SQ Sequence 1107 AA;
 Query Match 97.0%; Score 1149; DB 3; Length 1107;
 Best Local Similarity 93.6%; Pred. No. 7.1e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPPPPGSGRGGPFCADGVAGPKPAGERGSPGPA 48
 DB 297 GPPGAGEGKRGARGEPGTGLPPPPGSGRGGPFCADGVAGPKPAGERGSPGPA 356
 QY 49 GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPDGKTPGPPAGQDGRPGPPPGARGQA 108
 DB 357 GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPDGKTPGPPAGQDGRPGPPPGARGQA 416
 QY 109 GVMGFPKPKAAGEPKGAGERGVPPGAVGPKDGEAGAGQPPGPPGARGQA 168
 DB 417 GVMGFPKPKAAGEPKGAGERGVPPGAVGPKDGEAGAGQPPGPPGARGQA 476
 QY 169 GSPFGQLPGPAGPPGAGEAKPGEGQGVPGDLGAPGSPGAG 208
 DB 477 GSPFGQLPGPAGPPGAGEAKPGEGQGVPGDLGAPGSPGAG 516
 RESULT 13
 ADE87050
 ID ADE87050 standard; protein; 1161 AA.
 XX
 AC ADE87050;
 XX
 XX 29-JAN-2004 (first entry)
 DT
 DE Human pancreatic cell protein sequence SeqID510.
 XX
 XX neoplastic pancreatic cell; pancreatic cancer;
 KW cancer death; cytostatic; vaccine; gene therapy;
 XX

KW non-cancerous pancreas disease; human.
 XX Homo sapiens.
 XX WO2003060145-A2.
 XX 24-JUL-2003.
 XX 19-DEC-2002; 2002WO-US040655.
 XX 21-DEC-2001; 2001US-0342769P.
 XX (DIAD-) DIADEXUS INC.
 XX Sun Y, Liu C;
 XX WPI; 2003-587286/55.
 DR N-PSDB; ADE87387.
 XX
 PT New pancreatic specific nucleic acid molecule or protein for diagnosing,
 PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
 PT non-cancerous disease states of the pancreas.
 XX
 PS Claim 12; SEQ ID NO 510; 635pp; English.
 CC This invention relates to novel nucleic acids and proteins present in
 CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common
 CC cause of cancer death worldwide, therefore accurate methods of diagnosis
 CC and treatment are required. Compounds which modulate the proteins of the
 CC invention may have cytostatic activity and the protein and DNA sequences
 CC of the invention may be useful for the development of a vaccine or in
 CC gene therapy. The composition and methods are useful in diagnosing,
 CC staging, imaging, monitoring, preventing or treating pancreatic cancer
 CC and non-cancerous disease states of the pancreas. The present sequence is
 CC that of a human pancreatic protein of the invention.
 XX
 SQ Sequence 1161 AA;
 Query Match 97.0%; Score 1149; DB 7; Length 1161;
 Best Local Similarity 93.6%; Pred. No. 7.4e-67;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPPGPTGLPPPPGSGRGGPFCADGVAGPKPAGERGSPGPA 48
 DB 155 GPPGAGEGKRGARGEPGTGLPPPPGSGRGGPFCADGVAGPKPAGERGSPGPA 214
 QY 49 GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPDGKTPGPPAGQDGRPGPPPGARGQA 108
 DB 215 GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPDGKTPGPPAGQDGRPGPPPGARGQA 274
 QY 109 GVMGFPKPKAAGEPKGAGERGVPPGAVGPKDGEAGAGQPPGPPGARGQA 168
 DB 275 GVMGFPKPKAAGEPKGAGERGVPPGAVGPKDGEAGAGQPPGPPGARGQA 334
 QY 169 GSPFGQLPGPAGPPGAGEAKPGEGQGVPGDLGAPGSPGAG 208
 DB 335 GSPFGQLPGPAGPPGAGEAKPGEGQGVPGDLGAPGSPGAG 374
 RESULT 14
 AAR89469
 ID AAR89469 standard; protein; 1169 AA.
 XX
 AC AAR89469;
 XX
 XX 01-OCT-1996 (first entry)
 DT
 DE Collagen/BMP-2B fusion protein.
 XX
 XX Bone morphogenic protein 2B; BMP-2B; collagen 1A; osteogenesis;
 KW fusion protein.
 XX
 XX Synthetic.

```

XX FH Key Location/Qualifiers
XX DE 1. .1057
XX DE /label= Collagen-IA
XX DE /note= "collagen IA alpha-helical domain"
XX FT Misc-difference 887
XX FT /note= "unidentified amino acid"
XX FT Misc-difference 890
XX FT /note= "unidentified amino acid"
XX FT Peptide
XX FT /label= Linker_peptide
XX FT Domain
XX FT /label= BMP-2B
XX FT /note= "human mature BMP-2B"
XX CA2151547-A.
XX PN
XX PD 11-DEC-1995.
XX PP 12-JUN-1995; 95CA-02151547.
XX PR 10-JUN-1994; 94US-00259263.
XX PA (USSU ) US SURGICAL CORP.
XX PI Gruskin EA, Espino P;
XX DR WPI; 1996-140144/15.
XX DR N-PSDB; AAT16515.
XX PT Chimaeric DNA encoding protein contg. extracellular matrix protein domain
XX PT - and cellular regulatory factor domain, partic. useful as osteogenic
XX PT agents, also related vectors, transformed cells and chimaeric proteins.
XX PS Disclosure; Fig 5; 59pp; English.
XX CC A fusion protein (AAR89469) comprises the alpha-helical region of human
XX CC collagen I(a) linked to the human mature bone morphogenic protein 2B
XX CC (BMP2B). It can be expressed in Escherichia coli transformants carrying a
XX CC vector incorporating a chimeric gene (AAT16515) coding for the fusion.
XX CC The BMP moiety induces osteogenesis, while the collagen moiety provides
XX CC an integral substratum or scaffolding for the BMP and cells involved in
XX CC reconstruction and growth. The fusion protein provides sustained release
XX CC and delivery of BMP to a target tissue
XX SQ
Query Match 97.0%; Score 1149; DB 2; Length 1169;
Best Local Similarity 93.6%; Pred. No. 7.4e-67;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGPTGLPGPGERGSGRPFPGADGVAGPKPAGERGSGPGA 48
DB 297 GPPGAGEGRGARGEPGPTGLPGPGERGSGRPFPGADGVAGPKPAGERGSGPGA 356
QY 49 GPKSGPGEAGPGEAGLPGKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
DB 357 GPKSGPGEAGPGEAGLPGKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 416
QY 109 GVMGPPGPKGAAGEPKKAGRGVPFGAVGAGKDGAGAGQGGPPGPPGAGERGQCPA 168
DB 417 GVMGPPGPKGAAGEPKKAGRGVPFGAVGAGKDGAGAGQGGPPGPPGAGERGQCPA 476
QY 169 GSPGFQGLPGPAGPGEAGKPGEGQGVGDLAGAPGSPGAG 208
DB 477 GSPGFQGLPGPAGPGEAGKPGEGQGVGDLAGAPGSPGAG 516
RESULT 15
AAV84537
ID AAV84537 standard; protein; 1169 AA.
XX AC AAV84537;

```

```

XX 25-JUL-2000 (first entry)
XX Amino acid sequence of a chimeric collagen 1 (alpha1)/BMP-2B protein.
XX Extracellular matrix protein; self aggregation; hydroxylated proline;
XX trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
XX collagen; fibrinogen; fibronectin; post translational hydroxylation;
XX ss. bone morphogenic protein; BMP-2B; chimera.
XX Homo sapiens.
XX OS Unidentified.
XX OS Chimeric.
XX FH Key Location/Qualifiers
XX FT Misc-difference 677
XX FT /note= "Ala encoded by G"
XX FT Misc-difference 887
XX FT /note= "unspecified amino acid encoded by CT"
XX FT Misc-difference 890
XX FT /note= "unspecified amino acid encoded by CT"
XX EP992586-A2.
XX 12-APR-2000.
XX 07-OCT-1999; 99EP-00119184.
XX 09-OCT-1998; 98US-00169768.
XX (USSU ) US SURGICAL CORP.
XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX WPI; 2000-259138/23.
XX N-PSDB; AAL12497.
XX Production of extracellular matrix proteins containing 4-trans-
XX hydroxyproline results in native self aggregating proteins, useful on
XX medical implants.
XX Claim 22; Fig 13; 260pp; English.
XX The specification describes a method for producing an extracellular
XX matrix protein or its fragment. The extracellular matrix protein is
XX capable of self aggregating in a cell which does not ordinarily
XX hydroxylated prolines. The method comprises optimising a nucleic acid
XX sequence for expression in the cell by substitution of codons preferred
XX by that cell for naturally occurring codons not preferred by the cell;
XX incorporating the nucleic acid sequence into the cell; and contacting the
XX cell with a hypertonic growth medium containing at least one amino acid,
XX selected from the group consisting of trans-4-hydroxyproline and 3-
XX hydroxyproline to allow at least one of the amino acids to be assimilated
XX into the cell and incorporated into the extracellular matrix protein. The
XX method may be used to make host cells assimilate and incorporate trans-4-
XX hydroxyproline into proteins. This is especially useful in the
XX recombinant production of proteins such as collagen, fibrinogen and
XX fibronectin whose ability to self aggregate and produce functional
XX proteins depends on the post translational hydroxylation of proline. The
XX method is also useful in studying the structure and function of
XX polypeptides which do not normally contain trans-4-hydroxyproline. The
XX present sequence represents a chimeric collagen 1 (alpha1)/bone
XX morphogenic protein-2B (bmp-2b) protein, which may be produced using the
XX method of the invention
XX SQ Sequence 1169 AA;
Query Match 97.0%; Score 1149; DB 3; Length 1169;
Best Local Similarity 93.6%; Pred. No. 7.4e-67;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
QY 1 GPP-----GEPGPTGLPGPGERGSGRPFPGADGVAGPKPAGERGSGPGA 48

```

Db	297	GPPGACBEGKRGARGFPGTGLPGPPGERGGFCGRGFPQADGVAGKGPAGERGSPGPA	356
Qy	49	GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPDCKTGPAGQDGRPGPPGPPGARGQA	108
Db	357	GPKGSPGEAGRPGEAGLPGAKGLTSGSPGPDCKTGPAGQDGRPGPPGPPGARGQA	416
Qy	109	GVMGPPGPKGAAGEPPKAGRGVBPBGA VGPAGKDGEGAAGQPPGPPGPPGARGQA	168
Db	417	GVMGPPGPKGAAGEPPKAGRGVBPBGA VGPAGKDGEGAAGQPPGPPGPPGARGQA	476
Qy	169	GSPGFQGLPGPAGPPGEGAGKPGEGQGVPGDLGAPGSPGAG	208
Db	477	GSPGFQGLPGPAGPPGEGAGKPGEGQGVPGDLGAPGSPGAG	516

Search completed: June 17, 2005, 15:08:15
Job time : 37.1497 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:46:04 ; Search time 8.87129 Seconds
(without alignment)
1758.668 Million cell updates/sec

Title: US-10-658-989A-1
Perfect score: 1184
Sequence: 1 GPGCEPPTGLPGPPGREGG.....GEQVPGDLGNPGSPAGG 209

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1149	97.0	1341	3	US-08-963-825-18
2	1149	97.0	1341	3	US-09-500-811-18
3	1149	97.0	1341	3	US-09-570-573-18
4	1149	97.0	1341	3	US-09-548-608-18
5	1149	97.0	1461	4	US-09-585-887-9
6	1149	97.0	1461	4	US-09-289-578-9
7	1149	97.0	1464	4	US-09-331-347C-21
8	1120	94.6	1057	3	US-08-931-820-1
9	1116	94.3	595	3	US-09-213-849-48
10	1116	94.3	595	3	US-09-213-849-50
11	1116	94.3	822	3	US-09-213-849-49
12	908	76.7	492	4	US-08-468-996-12
13	866	73.1	1017	4	US-08-468-996-10
14	862	72.8	1060	3	US-08-931-820-3
15	862	72.8	1418	3	US-08-963-825-20
16	862	72.8	1418	3	US-09-010-999-1
17	862	72.8	1418	3	US-09-500-811-20
18	862	72.8	1418	3	US-09-570-573-20
19	862	72.8	1418	3	US-09-548-608-20
20	839	70.9	1442	2	US-08-316-650-12
21	839	70.9	1442	5	PCT-US95-02251-12
22	776	65.5	1057	3	US-08-931-820-4
23	765	64.6	492	4	US-08-468-996-11
24	755	63.8	1078	3	US-08-963-825-21
25	755	63.8	1078	3	US-09-500-811-21
26	755	63.8	1078	3	US-09-570-573-21
27	755	63.8	1078	3	US-09-548-608-21

28 745 62.9 1366 4 US-09-585-887-10 Sequence 10, Appl
29 745 62.9 1366 4 US-09-289-578-10 Sequence 10, Appl
30 745 62.9 1366 4 US-09-949-016-5882 Sequence 5882, Ap
31 744 62.8 1024 3 US-08-931-820-2 Sequence 2, Appl
32 744 62.8 1366 3 US-08-963-825-19 Sequence 19, Appl
33 744 62.8 1366 3 US-09-500-811-19 Sequence 19, Appl
34 744 62.8 1366 3 US-09-570-573-19 Sequence 19, Appl
35 744 62.8 1366 3 US-09-548-608-19 Sequence 19, Appl
36 657 55.5 330 1 US-08-642-255-32 Sequence 32, Appl
37 657 55.5 408 1 US-07-609-716-65 Sequence 65, Appl
38 657 55.5 408 3 US-08-475-411A-65 Sequence 65, Appl
39 657 55.5 408 3 US-08-478-029A-65 Sequence 3, Appl
40 655.5 55.4 504 3 US-09-219-849-3 Sequence 52, Appl
41 655.5 55.4 561 1 US-08-642-255-52 Sequence 51, Appl
42 650.5 54.9 234 1 US-08-642-255-51 Sequence 4, Appl
43 648 54.7 720 3 US-09-219-849-4 Sequence 53, Appl
44 648 54.7 777 1 US-08-642-255-53 Sequence 62, Appl
45 646.5 54.6 1064 1 US-08-642-255-62

ALIGNMENTS

RESULT 1

US-08-963-825-18

; Sequence 18, Application US/08963825

; Patent No. 6110689

; GENERAL INFORMATION:

; APPLICANT: Qvist, Per

; APPLICANT: Bonde, Martin

; TITLE OF INVENTION: A Method for Assaying Collagen Fragments

; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the

; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

; TITLE OF INVENTION: Disorders Associated with the Metabolism of

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Darby & Darby PC

; STREET: 805 Third Avenue

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/963,825

; FILING DATE:

; CLASSIFICATION: 436

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/187,319

; FILING DATE: 21-JAN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Gogoris, Adda C

; REGISTRATION NUMBER: 29,714

; REFERENCE/DOCKET NUMBER: 4305/08701

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-527-7700

; TELEFAX: 212-753-6237

; TELEX: 236687

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1341 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; IMMEDIATE SOURCE:

; CLONE: COLLAGEN ALPHA 1 (I)

US-08-963-825-18

Query Match 97.0%; Score 1149; DB 3; Length 1341;
 Best Local Similarity 93.6%; Pred. No. 2.9e-72;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGGRGGSGRFPAGDGVAGPKGAGRGSPGPA 48
 |||||
 Db 334 GPPGAGGEGKRGARGGPGTGLPGPPGGRGGSGRFPAGDGVAGPKGAGRGSPGPA 393

QY 49 GPKGSPGAGRPAGLPGAKGLTGTSPGSPGPKGTGPPGAGDGRPPGPPGARGOA 108
 |||||
 Db 394 GPKGSPGAGRPAGLPGAKGLTGTSPGSPGPKGTGPPGAGDGRPPGPPGARGOA 453

QY 109 GVMGFPKGAAGEPKAGRGVPGPAGVGPAGKDGAGAGQPPGPPGARGQGPA 168
 |||||
 Db 454 GVMGFPKGAAGEPKAGRGVPGPAGVGPAGKDGAGAGQPPGPPGARGQGPA 513

QY 169 GSPGFQGLPGPAGPGEAGKFGEGVPGDGLGAPGSPGAG 208
 |||||
 Db 514 GSPGFQGLPGPAGPGEAGKFGEGVPGDGLGAPGSPGAG 553

RESULT 3
 US-09-500-811-18
 ; Sequence 18, Application US/09500811
 ; Patent No. 6323314
 ; GENERAL INFORMATION:
 ; APPLICANT: Qvist, Per
 ; APPLICANT: Bonde, Martin
 ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 ; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
 ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
 ; TITLE OF INVENTION: Disorders Associated with the Metabolism of
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby PC
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/500,811
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/187,319
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gogoris, Adda C
 ; REGISTRATION NUMBER: 29,714
 ; REFERENCE/DOCKET NUMBER: 4305/08701
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-527-7700
 ; TELEFAX: 212-753-6237
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1341 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; IMMEDIATE SOURCE:
 ; CLONE: COLLAGEN ALPHA 1 (I)
 ; US-09-500-811-18

Query Match 97.0%; Score 1149; DB 3; Length 1341;
 Best Local Similarity 93.6%; Pred. No. 2.9e-72;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGGRGGSGRFPAGDGVAGPKGAGRGSPGPA 48
 |||||
 Db 334 GPPGAGGEGKRGARGGPGTGLPGPPGGRGGSGRFPAGDGVAGPKGAGRGSPGPA 393

QY 49 GPKGSPGAGRPAGLPGAKGLTGTSPGSPGPKGTGPPGAGDGRPPGPPGARGOA 108
 |||||
 Db 394 GPKGSPGAGRPAGLPGAKGLTGTSPGSPGPKGTGPPGAGDGRPPGPPGARGOA 453

QY 109 GVMGFPKGAAGEPKAGRGVPGPAGVGPAGKDGAGAGQPPGPPGARGQGPA 168
 |||||
 Db 454 GVMGFPKGAAGEPKAGRGVPGPAGVGPAGKDGAGAGQPPGPPGARGQGPA 513

QY 169 GSPGFQGLPGPAGPGEAGKFGEGVPGDGLGAPGSPGAG 208
 |||||
 Db 514 GSPGFQGLPGPAGPGEAGKFGEGVPGDGLGAPGSPGAG 553

RESULT 3
 US-09-570-573-18
 ; Sequence 18, Application US/09570573
 ; Patent No. 6342361
 ; GENERAL INFORMATION:
 ; APPLICANT: Qvist, Per
 ; APPLICANT: Bonde, Martin
 ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 ; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
 ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
 ; TITLE OF INVENTION: Disorders Associated with the Metabolism of
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby PC
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/570,573
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/187,319
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gogoris, Adda C
 ; REGISTRATION NUMBER: 29,714
 ; REFERENCE/DOCKET NUMBER: 4305/08701
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-527-7700
 ; TELEFAX: 212-753-6237
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1341 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; IMMEDIATE SOURCE:
 ; CLONE: COLLAGEN ALPHA 1 (I)
 ; US-09-570-573-18

Query Match 97.0%; Score 1149; DB 3; Length 1341;

Best Local Similarity 93.6%; Pred. No. 2.9e-72; Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPPTGLPPGGERGGPSRGGPGADGVAGPKGPAGERGSPGA 48
 DB 334 GPPGAGEGKRGARGEPGPTGLPPGGERGGPSRGGPGADGVAGPKGPAGERGSPGA 393
 QY 49 GPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
 DB 394 GPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 453
 QY 109 GVMGFPKGAAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 168
 DB 454 GVMGFPKGAAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 513
 QY 169 GSPGFQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAG 208
 DB 514 GSPGFQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGARG 553

RESULT 4
 US-09-548-608-18
 ; Sequence 18, Application US/09548608
 ; Patent No. 6355442
 ; GENERAL INFORMATION:
 ; APPLICANT: Qvist, Per
 ; APPLICANT: Bonde, Martin
 ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 ; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
 ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
 ; TITLE OF INVENTION: Disorders Associated with the Metabolism of
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby PC
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/548,608
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/187,319
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Cogoris, Adda C
 ; REGISTRATION NUMBER: 29,714
 ; REFERENCE/DOCKET NUMBER: 4305/08701
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-527-7700
 ; TELEFAX: 212-753-6237
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1341 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; IMMEDIATE SOURCE:
 ; CLONE: COLLAGEN ALPHA 1 (I)
 US-09-548-608-18

Query Match 97.0%; Score 1149; DB 3; Length 1341;
 Best Local Similarity 93.6%; Pred. No. 2.9e-72;

Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPPTGLPPGGERGGPSRGGPGADGVAGPKGPAGERGSPGA 48
 DB 334 GPPGAGEGKRGARGEPGPTGLPPGGERGGPSRGGPGADGVAGPKGPAGERGSPGA 393
 QY 49 GPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
 DB 394 GPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 453
 QY 109 GVMGFPKGAAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 168
 DB 454 GVMGFPKGAAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 513
 QY 169 GSPGFQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAG 208
 DB 514 GSPGFQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGARG 553

RESULT 5
 US-09-585-887-9
 ; Sequence 9, Application US/09585887
 ; Patent No. 6413742
 ; GENERAL INFORMATION:
 ; APPLICANT: Olsen, David R
 ; APPLICANT: Chang, Robert
 ; APPLICANT: McMullin, Hugh
 ; APPLICANT: Hitzeman, Ronald A.
 ; APPLICANT: Chisholm, George
 ; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
 ; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
 ; TITLE OF INVENTION: CELLS
 ; FILE REFERENCE: 225002030400
 ; CURRENT APPLICATION NUMBER: US/09/585,887
 ; CURRENT FILING DATE: 2000-05-31
 ; PRIOR APPLICATION NUMBER: 09/289,578
 ; PRIOR FILING DATE: 1999-04-09
 ; PRIOR APPLICATION NUMBER: 60/084,828
 ; PRIOR FILING DATE: 1998-05-08
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 9
 ; LENGTH: 1461
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-585-887-9

Query Match 97.0%; Score 1149; DB 4; Length 1461;
 Best Local Similarity 93.6%; Pred. No. 3.1e-72;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPPTGLPPGGERGGPSRGGPGADGVAGPKGPAGERGSPGA 48
 DB 455 GPPGAGEGKRGARGEPGPTGLPPGGERGGPSRGGPGADGVAGPKGPAGERGSPGA 514
 QY 49 GPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
 DB 515 GPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 574
 QY 109 GVMGFPKGAAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 168
 DB 575 GVMGFPKGAAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 634
 QY 169 GSPGFQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGAG 208
 DB 635 GSPGFQGLPGPAGPGEAGKPGEGQVPGDLGAPGSPGARG 674

RESULT 6
 US-09-289-578-9
 ; Sequence 9, Application US/09289578
 ; Patent No. 6428978
 ; GENERAL INFORMATION:

; APPLICANT: DE WOLF, FREDERIK A.
 ; APPLICANT: MOOREK, ANDREAS
 ; APPLICANT: WERTEN, MARC W.T.
 ; APPLICANT: WIND, RICHEL D.
 ; APPLICANT: VAN DEN BOSCH, TANJA J.
 ; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
 ; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
 ; TITLE OF INVENTION: PREPARATION THEREOF
 ; FILE REFERENCE: 2728-2
 ; CURRENT APPLICATION NUMBER: US/09/219,849
 ; CURRENT FILING DATE: 1998-12-23
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 48
 ; LENGTH: 595
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; OTHER INFORMATION: amino acid sequence
 US-09-219-849-48

Query Match 94.3%; Score 1116; DB 3; Length 595;
 Best Local Similarity 90.5%; Pred. No. 2.7e-70;
 Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

Qy 1 GPP-----GEPGPTGLPGERGSGRGGPAGDVGAGPKGPGAGERSGPPA 48
 Db 279 GPPGAGEGKRGARGEPGSGLPGERGGSGRGGPAGDVGAGPKGPGAGERSGPPA 338
 Qy 49 GPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGODGRRPGRPPGARGQA 108
 Db 339 GPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGODGRRPGRPPGARGQA 398
 Qy 109 GVMGFPKGAAGEPGKAGRGVPGPGAVGPGAGKDGGAQAQGGPAGPAGERGEOGPA 168
 Db 399 GVMGFPKGTAGEPGKAGRGVPGPGAVGPGAGKDGGAQAQGGPAGPAGERGEOGPA 498
 Qy 169 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 208
 Db 459 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 498

RESULT 10
 US-09-219-849-50
 ; Sequence 50, Application US/09219849
 ; Patent No. 6150081
 ; GENERAL INFORMATION:
 ; APPLICANT: VAN HEERDE, GEORGE V.
 ; APPLICANT: VAN RIJN, ALEXIS C.
 ; APPLICANT: BOWMSTRA, JAN B.
 ; APPLICANT: DE WOLF, FREDERIK A.
 ; APPLICANT: MOOREK, ANDREAS
 ; APPLICANT: WERTEN, MARC W.T.
 ; APPLICANT: WIND, RICHEL D.
 ; APPLICANT: VAN DEN BOSCH, TANJA J.
 ; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
 ; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
 ; TITLE OF INVENTION: PREPARATION THEREOF
 ; FILE REFERENCE: 2728-2
 ; CURRENT APPLICATION NUMBER: US/09/219,849
 ; CURRENT FILING DATE: 1998-12-23
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 49
 ; LENGTH: 822
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; OTHER INFORMATION: amino acid sequence
 US-09-219-849-49

Query Match 94.3%; Score 1116; DB 3; Length 822;
 Best Local Similarity 90.5%; Pred. No. 3.6e-70;
 Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

Qy 1 GPP-----GEPGPTGLPGERGSGRGGPAGDVGAGPKGPGAGERSGPPA 48
 Db 279 GPPGAGEGKRGARGEPGSGLPGERGGSGRGGPAGDVGAGPKGPGAGERSGPPA 338
 Qy 49 GPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGODGRRPGRPPGARGQA 108
 Db 339 GPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGODGRRPGRPPGARGQA 398
 Qy 109 GVMGFPKGAAGEPGKAGRGVPGPGAVGPGAGKDGGAQAQGGPAGPAGERGEOGPA 168
 Db 399 GVMGFPKGTAGEPGKAGRGVPGPGAVGPGAGKDGGAQAQGGPAGPAGERGEOGPA 498
 Qy 169 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 208
 Db 459 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 498

RESULT 12
 US-09-219-849-50
 ; Sequence 50, Application US/09219849
 ; Patent No. 6150081
 ; GENERAL INFORMATION:
 ; APPLICANT: VAN HEERDE, GEORGE V.
 ; APPLICANT: VAN RIJN, ALEXIS C.
 ; APPLICANT: BOWMSTRA, JAN B.
 ; APPLICANT: DE WOLF, FREDERIK A.
 ; APPLICANT: MOOREK, ANDREAS
 ; APPLICANT: WERTEN, MARC W.T.
 ; APPLICANT: WIND, RICHEL D.
 ; APPLICANT: VAN DEN BOSCH, TANJA J.
 ; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
 ; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
 ; TITLE OF INVENTION: PREPARATION THEREOF
 ; FILE REFERENCE: 2728-2
 ; CURRENT APPLICATION NUMBER: US/09/219,849
 ; CURRENT FILING DATE: 1998-12-23
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 50
 ; LENGTH: 595
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; OTHER INFORMATION: amino acid sequence
 US-09-219-849-50

Query Match 94.3%; Score 1116; DB 3; Length 595;
 Best Local Similarity 90.5%; Pred. No. 2.7e-70;
 Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

Qy 1 GPP-----GEPGPTGLPGERGSGRGGPAGDVGAGPKGPGAGERSGPPA 48
 Db 279 GPPGAGEGKRGARGEPGSGLPGERGGSGRGGPAGDVGAGPKGPGAGERSGPPA 338
 Qy 49 GPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGODGRRPGRPPGARGQA 108
 Db 339 GPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGODGRRPGRPPGARGQA 398
 Qy 109 GVMGFPKGAAGEPGKAGRGVPGPGAVGPGAGKDGGAQAQGGPAGPAGERGEOGPA 168
 Db 399 GVMGFPKGTAGEPGKAGRGVPGPGAVGPGAGKDGGAQAQGGPAGPAGERGEOGPA 498
 Qy 169 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 208
 Db 459 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 498

RESULT 11
 US-09-219-849-49
 ; Sequence 49, Application US/09219849
 ; Patent No. 6150081
 ; GENERAL INFORMATION:
 ; APPLICANT: VAN HEERDE, GEORGE V.
 ; APPLICANT: VAN RIJN, ALEXIS C.
 ; APPLICANT: BOWMSTRA, JAN B.
 ; APPLICANT: DE WOLF, FREDERIK A.
 ; APPLICANT: MOOREK, ANDREAS
 ; APPLICANT: WERTEN, MARC W.T.
 ; APPLICANT: WIND, RICHEL D.
 ; APPLICANT: VAN DEN BOSCH, TANJA J.
 ; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
 ; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
 ; TITLE OF INVENTION: PREPARATION THEREOF
 ; FILE REFERENCE: 2728-2
 ; CURRENT APPLICATION NUMBER: US/09/219,849
 ; CURRENT FILING DATE: 1998-12-23
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 49
 ; LENGTH: 822
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; OTHER INFORMATION: amino acid sequence
 US-09-219-849-49

Query Match 94.3%; Score 1116; DB 3; Length 822;
 Best Local Similarity 90.5%; Pred. No. 3.6e-70;
 Matches 199; Conservative 4; Mismatches 5; Indels 12; Gaps 1;

Qy 1 GPP-----GEPGPTGLPGERGSGRGGPAGDVGAGPKGPGAGERSGPPA 48
 Db 279 GPPGAGEGKRGARGEPGSGLPGERGGSGRGGPAGDVGAGPKGPGAGERSGPPA 338
 Qy 49 GPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGODGRRPGRPPGARGQA 108
 Db 339 GPKSGPGEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGODGRRPGRPPGARGQA 398
 Qy 109 GVMGFPKGAAGEPGKAGRGVPGPGAVGPGAGKDGGAQAQGGPAGPAGERGEOGPA 168
 Db 399 GVMGFPKGTAGEPGKAGRGVPGPGAVGPGAGKDGGAQAQGGPAGPAGERGEOGPA 498
 Qy 169 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 208
 Db 459 GSPGFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAG 498

RESULT 12
 US-09-219-849-50
 ; Sequence 50, Application US/09219849
 ; Patent No. 6150081
 ; GENERAL INFORMATION:
 ; APPLICANT: VAN HEERDE, GEORGE V.
 ; APPLICANT: VAN RIJN, ALEXIS C.
 ; APPLICANT: BOWMSTRA, JAN B.
 ; APPLICANT: DE WOLF, FREDERIK A.
 ; APPLICANT: MOOREK, ANDREAS
 ; APPLICANT: WERTEN, MARC W.T.
 ; APPLICANT: WIND, RICHEL D.
 ; APPLICANT: VAN DEN BOSCH, TANJA J.
 ; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
 ; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
 ; TITLE OF INVENTION: PREPARATION THEREOF
 ; FILE REFERENCE: 2728-2
 ; CURRENT APPLICATION NUMBER: US/09/219,849
 ; CURRENT FILING DATE: 1998-12-23
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 50
 ; LENGTH: 595
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 ; OTHER INFORMATION: amino acid sequence
 US-09-219-849-50

```

US-08-468-996-12
; Sequence 12, Application US/08468996
; Patent No. 6645504
; GENERAL INFORMATION:
; APPLICANT: Weiner, Howard
; APPLICANT: Miller, Ariel
; APPLICANT: Zheng, Zheng
; APPLICANT: Ahmad, Al-Sabbagh
; TITLE OF INVENTION: BYSTANDER SUPPRESSION OF TYPE 1 DIABETES BY ORAL ADMINISTRATION
; FILE REFERENCE: 1010/16959-US3
; CURRENT APPLICATION NUMBER: US/08/468,996
; CURRENT FILING DATE: 2003-02-07
; PRIOR FILING DATE: 2003-02-07
; PRIOR FILING DATE: 1992-02-28
; PRIOR FILING DATE: 1990-02-21
; PRIOR FILING DATE: 1990-02-21
; PRIOR FILING DATE: 1990-10-15
; PRIOR FILING DATE: 1987-06-24
; PRIOR FILING DATE: 1989-12-20
; PRIOR FILING DATE: 1989-12-20
; PRIOR FILING DATE: 1990-03-02
; PRIOR FILING DATE: 1990-07-10
; PRIOR FILING DATE: 1990-07-14
; PRIOR FILING DATE: 1989-07-14
; PRIOR FILING DATE: 1990-10-31
; PRIOR FILING DATE: 1990-10-10
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 12
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Bos taurus
US-08-468-996-12

Query Match 76.7%; Score 908; DB 4; Length 492;
Best Local Similarity 76.0%; Pred. No. 6e-56;
Matches 165; Conservative 9; Mismatches 25; Indels 18; Gaps 2;

QY 1 GPP-----GEPDTGLPPGRRGPGGPGGADGVAGPKGAPGERSGPGPA 48
Db 280 GPPGAGEGKRGARGEPGDTGLPPGRRGPGGPGGADGVAGPKGAPGERSGPGPA 339
QY 49 GPKGSPGAGRPCEAGLPCAKGLTGSPGSPDGKGTGPPGAGODGRGPPGPGARQQA 108
Db 340 GPKGSPGAGRPCEAGLPCAKGLTGSPGSPDGKGTGPPGAGQNGRPGPPGPGARQQA 399
QY 109 GVMGPPGPKGAAGEPGKAGRGVPPGPGAVGAGPKDGEAGAGQPPGAPGAGEGQGPA 168
Db 400 GVMGPPGPKGAAGEPGKAGRGVPPG-----GNDGAKGDAGAGLPGPKGDRDAGPK 453
QY 169 GSPGQCLPGDAGPGEAGKPGEQGVGDLGAPGSG 205
Db 454 GADGAPGAPGDKGEAGPSQPGKAGKPGDAGAKGDAG 490

RESULT 13
US-08-468-996-10
; Sequence 10, Application US/08468996
; Patent No. 6645504
; GENERAL INFORMATION:
; APPLICANT: Weiner, Howard
; APPLICANT: Miller, Ariel
; APPLICANT: Zheng, Zheng
; APPLICANT: Ahmad, Al-Sabbagh
; TITLE OF INVENTION: BYSTANDER SUPPRESSION OF TYPE 1 DIABETES BY ORAL ADMINISTRATION
; FILE REFERENCE: 1010/16959-US3
; CURRENT APPLICATION NUMBER: US/08/468,996
; CURRENT FILING DATE: 2003-02-07
; PRIOR FILING DATE: 2003-02-07
; PRIOR FILING DATE: 1992-02-28
; PRIOR FILING DATE: 1990-02-21
; PRIOR FILING DATE: 1990-02-21
; PRIOR FILING DATE: 1990-10-15
; PRIOR FILING DATE: 1987-06-24
; PRIOR FILING DATE: 1989-12-20
; PRIOR FILING DATE: 1989-12-20
; PRIOR FILING DATE: 1990-03-02
; PRIOR FILING DATE: 1990-07-10
; PRIOR FILING DATE: 1990-07-14
; PRIOR FILING DATE: 1989-07-14
; PRIOR FILING DATE: 1990-10-31
; PRIOR FILING DATE: 1990-10-10
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 1017
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-468-996-10

Query Match 73.1%; Score 866; DB 4; Length 1017;
Best Local Similarity 74.5%; Pred. No. 9.3e-53;
Matches 155; Conservative 11; Mismatches 42; Indels 0; Gaps 0;

QY 1 GPPGEPGTGLPPGRRGPGGPGGADGVAGPKGAPGERSGPGAGPKGSPGAGRP 60
Db 292 GARGEPGVGPIGPPGRRGAPGNRGGFGQDLGAGPKGAPGERSGGLAGPKGANGDPGRP 351
QY 61 GEAGLPGAKGLTGSPGSPGDPGKGTGPPGAGODGRGPPGPGARGAGVGMGPPGKGA 120
Db 352 GEPGLPGARGLTGPPGAGQKVGPSGAGDGRGPPGPGQARGQGVGMGPPGPKGAN 411
QY 121 GEPKAGRGVPPGPGAVGAGPKDGEAGAGQPPGAPGAGRGQGPAGSPGFGGLFPGA 180
Db 412 GEPKAGRGVPPGPGAVGAGPKDGEAGAGQPPGAPGAGRGQGPAGSPGFGGLFPGA 471
QY 181 GPPGAGKPGEQGVGPDGLGAPGSGPAG 208
Db 472 GPPGAGKPGDQGVGPGAGAPGLVGPAG 499

RESULT 14
US-08-931-820-3
; Sequence 3, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 3:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type II
US-08-931-820-3

```

Query Match 72.8%; Score 862; DB 3; Length 1060;
 Best Local Similarity 74.0%; Pred. No. 1.8e-52;
 Matches 154; Conservative 11; Mismatches 43; Indels 0; Gaps 0;

```

Qy 1 GPPGEPGPTGLPDPGPGGSGRFPAGDVGAGPKGAPAGRGSPGAPGKSPGCEAGRP 60
Db 311 GARGEPGVGPIGPPGERGAFGNRFPQDGLAGPKGAPGERGSLAGPKGANGDPGRP 370
Qy 61 GEAGLPKAGLGTSGSPGPGDKTTPPGAGGODGRPPGPPGARGQAGVGMGFPKPGAA 120
Db 371 GEPGLPGARGLTGRPDAGPKQKVGPSGAPGEDGRPPGPGARGQPGVGMGFPKPGAN 430
Qy 121 GEPKAGRGVPGPGCAVGPAGKDEAGACQAGPPGAPGAPGERGQGPAGSPGFOGLPGPA 180
Db 431 GEPKAGRGVPGPGCAVGPAGKDEAGACQAGPPGAPGAPGERGQGPAGSPGFOGLPGPP 490
Qy 181 GPPGAGKRGPGQGVPGDLGAPGSPGAP 208
Db 491 GPPGEGGKPGDQGVPGEGAGAPGLVGRPG 518

```

RESULT 15

US-08-963-825-20
 ; Sequence 20, Application US/08963825
 ; Patent No. 6110689

GENERAL INFORMATION:

APPLICANT: Qvist, Per
 TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
 TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
 TITLE OF INVENTION: Disorders Associated with the Metabolism of

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSER: Darby & Darby PC
 STREET: 805 Third Avenue
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/963,825

FILING DATE:

CLASSIFICATION: 436
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/187,319
 FILING DATE: 21-JAN-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Gogoris, Adda C
 REGISTRATION NUMBER: 29,714
 REFERENCE/DOCKET NUMBER: 4305/08701
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-527-7700
 TELEFAX: 212-753-6237

INFORMATION FOR SEQ ID NO: 20:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
US-08-963-825-20

```

Query Match 72.8%; Score 862; DB 3; Length 1418;
 Best Local Similarity 74.0%; Pred. No. 2.4e-52;
 Matches 154; Conservative 11; Mismatches 43; Indels 0; Gaps 0;

```

Qy 1 GPPGEPGPTGLPDPGPGGSGRFPAGDVGAGPKGAPAGRGSPGAPGKSPGCEAGRP 60
Db 423 GARGEPGVGPIGPPGERGAFGNRFPQDGLAGPKGAPGERGSLAGPKGANGDPGRP 482
Qy 61 GEAGLPKAGLGTSGSPGPGDKTTPPGAGGODGRPPGPPGARGQAGVGMGFPKPGAA 120
Db 483 GEPGLPGARGLTGRPDAGPKQKVGPSGAPGEDGRPPGPGARGQPGVGMGFPKPGAN 542
Qy 121 GEPKAGRGVPGPGCAVGPAGKDEAGACQAGPPGAPGAPGERGQGPAGSPGFOGLPGPA 180
Db 543 GEPKAGRGVPGPGCAVGPAGKDEAGACQAGPPGAPGAPGERGQGPAGSPGFOGLPGPP 602
Qy 181 GPPGAGKRGPGQGVPGDLGAPGSPGAP 208
Db 603 GPPGEGGKPGDQGVPGEGAGAPGLVGRPG 630

```

Search completed: June 17, 2005, 15:16:48
 Job time : 9.87129 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 15:02:42 ; Search time 22.8937 Seconds
(without alignments)
3505.413 Million cell updates/sec

Title: US-10-658-989A-1

Perfect score: 1184

Sequence: 1 GPPCEPPTGLPGRGGRGG.....GEGVPGDLGAPSPGAGG 209

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1714042 seqs, 38379560 residues

Total number of hits satisfying chosen parameters: 1714042

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1184	100.0	209	17	US-10-658-989A-1
2	1178	99.5	617	17	US-10-658-989A-2
3	1178	99.5	821	17	US-10-658-989A-3
4	1149	97.0	501	14	US-10-232-175-17
5	1149	97.0	1014	17	US-10-901-816A-5
6	1149	97.0	1014	17	US-10-901-816A-6
7	1149	97.0	1014	17	US-10-901-816A-9
8	1149	97.0	1014	17	US-10-901-816A-10
9	1149	97.0	1014	17	US-10-901-816A-11
10	1149	97.0	1057	15	US-10-104-889-16
11	1149	97.0	1057	15	US-10-104-889-20

12	1149	97.0	1107	15	US-10-104-889-11	Sequence 11, Appl
13	1149	97.0	1169	15	US-10-104-889-6	Sequence 8, Appl
14	1149	97.0	1171	15	US-10-104-889-8	Sequence 6, Appl
15	1149	97.0	1341	14	US-10-058-124-18	Sequence 18, Appl
16	1149	97.0	1388	15	US-10-104-889-10	Sequence 10, Appl
17	1149	97.0	1461	16	US-10-468-091-25	Sequence 25, Appl
18	1149	97.0	1464	10	US-09-918-715-261	Sequence 261, App
19	1149	97.0	1464	14	US-10-060-036-159	Sequence 159, App
20	1149	97.0	1464	14	US-10-171-311-36	Sequence 36, Appl
21	1149	97.0	1464	14	US-10-216-705-21	Sequence 21, Appl
22	1149	97.0	1464	14	US-10-149-352-2	Sequence 2, Appl
23	1149	97.0	1464	14	US-10-177-293-65	Sequence 65, Appl
24	1149	97.0	1464	14	US-10-301-822-28	Sequence 28, Appl
25	1149	97.0	1464	15	US-10-291-265-243	Sequence 243, App
26	1149	97.0	1464	16	US-10-357-851-1	Sequence 1, Appl
27	1149	97.0	1464	16	US-10-358-024-1	Sequence 1, Appl
28	1149	97.0	1464	16	US-10-734-564-79	Sequence 79, Appl
29	1149	97.0	1464	16	US-10-788-792-150	Sequence 150, App
30	1149	97.0	1464	16	US-10-474-794-261	Sequence 261, App
31	1149	97.0	1464	16	US-10-723-860-2289	Sequence 2289, Ap
32	1149	97.0	1464	17	US-10-852-335A-157	Sequence 157, App
33	1144	96.6	1449	15	US-10-402-089-8	Sequence 8, Appl
34	1144	96.6	1449	15	US-10-402-072A-8	Sequence 8, Appl
35	1141	96.4	1463	15	US-10-402-089-2	Sequence 2, Appl
36	1141	96.4	1463	15	US-10-402-072A-2	Sequence 2, Appl
37	1136	95.9	1014	17	US-10-901-816A-7	Sequence 7, Appl
38	1136	95.9	1014	17	US-10-901-816A-8	Sequence 8, Appl
39	1136	95.9	1014	17	US-10-901-816A-13	Sequence 13, Appl
40	1131	95.5	544	17	US-10-658-989A-4	Sequence 4, Appl
41	1116	94.3	595	15	US-10-342-331-48	Sequence 48, Appl
42	1116	94.3	595	15	US-10-342-331-50	Sequence 50, Appl
43	1116	94.3	822	15	US-10-342-331-49	Sequence 49, Appl
44	1116	94.3	1453	16	US-10-468-091-26	Sequence 26, Appl
45	1104	93.2	1014	17	US-10-901-816A-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1

; Sequence 1, Application US/10658989A
; Publication No. US20050101531A1
; GENERAL INFORMATION:
; APPLICANT: BOWSTRA, Jan Bastiaan
; APPLICANT: YUZO, Toda
; TITLE OF INVENTION: Use of recombinant gelatin-like proteins as plasma expanders and
; FILE REFERENCE: BOWSTRA-3
; CURRENT APPLICATION NUMBER: US/10/658,989A
; CURRENT FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: EP 02078745.3
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hu-1
US-10-658-989A-1

Query Match	100.0%	Score 1184;	DB 17;	Length 209;
Best Local Similarity	100.0%;	Pred. No. 7.9e-59;		
Matches: 209;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	GPPCEPPTGLPGRGGRGGPFPCADGACGKGPAGRGSPGAPKGSFGACGRP	60	
Db	1	GPPCEPPTGLPGRGGRGGPFPCADGACGKGPAGRGSPGAPKGSFGACGRP	60	
Qy	61	GEAGLPGAKGLTGSFGSPGPPGCKTGPQGPACQDGRPPGPPGARGQGVGFFGPKAA	120	

```

Db      61  GEAGLPKAGKLTGSPGSGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGA 120
QY      121  GEPKAGRGVPPGPGAVGAGKDGAGAGGPGGPGAGRGEGQGPAGSPGQGLPGPA 180
Db      121  GEPKAGRGVPPGPGAVGAGKDGAGAGGPGGPGAGRGEGQGPAGSPGQGLPGPA 180
QY      181  GPPGAGKPGQGVPPGDLGAPGSPGAG 209
Db      181  GPPGAGKPGQGVPPGDLGAPGSPGAG 209

RESULT 2
US-10-658-989A-2
; Sequence 2, Application US/10658989A
; Publication No. US20050101531A1
; GENERAL INFORMATION:
; APPLICANT: BOWSTRA, Jan Bastiaan
; APPLICANT: YUZO, Toda
; TITLE OF INVENTION: Use of recombinant gelatin-like proteins as plasma expanders and
; FILE REFERENCE: BOWSTRA-3
; CURRENT APPLICATION NUMBER: US/10/658,989A
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: EP 02078745.3
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hu-3
US-10-658-989A-2

Query Match      99.5%; Score 1178; DB 17; Length 617;
Best Local Similarity 100.0%; Pred. No. 3.9e-58;
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GPPGEPGTGLPGPPGGRGSGRFPDAGDVAGPKGPGAGRGSPGAGPKGSPGAGRP 60
Db      1  GPPGEPGTGLPGPPGGRGSGRFPDAGDVAGPKGPGAGRGSPGAGPKGSPGAGRP 60
QY      61  GEAGLPKAGKLTGSPGSGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGA 120
Db      61  GEAGLPKAGKLTGSPGSGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGA 120
QY      121  GEPKAGRGVPPGPGAVGAGKDGAGAGGPGGPGAGRGEGQGPAGSPGQGLPGPA 180
Db      121  GEPKAGRGVPPGPGAVGAGKDGAGAGGPGGPGAGRGEGQGPAGSPGQGLPGPA 180
QY      181  GPPGAGKPGQGVPPGDLGAPGSPGAG 208
Db      181  GPPGAGKPGQGVPPGDLGAPGSPGAG 208

US-10-658-989A-3
; Sequence 3, Application US/10658989A
; Publication No. US20050101531A1
; GENERAL INFORMATION:
; APPLICANT: BOWSTRA, Jan Bastiaan
; APPLICANT: YUZO, Toda
; TITLE OF INVENTION: Use of recombinant gelatin-like proteins as plasma expanders and
; FILE REFERENCE: BOWSTRA-3
; CURRENT APPLICATION NUMBER: US/10/658,989A
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: EP 02078745.3
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hu-3
US-10-658-989A-3

Query Match      99.5%; Score 1178; DB 17; Length 617;
Best Local Similarity 100.0%; Pred. No. 3.9e-58;
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GPPGEPGTGLPGPPGGRGSGRFPDAGDVAGPKGPGAGRGSPGAGPKGSPGAGRP 60
Db      1  GPPGEPGTGLPGPPGGRGSGRFPDAGDVAGPKGPGAGRGSPGAGPKGSPGAGRP 60
QY      61  GEAGLPKAGKLTGSPGSGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGA 120
Db      61  GEAGLPKAGKLTGSPGSGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGA 120
QY      121  GEPKAGRGVPPGPGAVGAGKDGAGAGGPGGPGAGRGEGQGPAGSPGQGLPGPA 180
Db      121  GEPKAGRGVPPGPGAVGAGKDGAGAGGPGGPGAGRGEGQGPAGSPGQGLPGPA 180
QY      181  GPPGAGKPGQGVPPGDLGAPGSPGAG 208
Db      181  GPPGAGKPGQGVPPGDLGAPGSPGAG 208

US-10-658-989A-4
; Sequence 4, Application US/10232175
; Publication No. US20030064074A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Kivirikko, Kari I.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Olsen, David R.
; APPLICANT: Polarek, James W.
; TITLE OF INVENTION: RECOMBINANT GELATINS IN VACCINES
; FILE REFERENCE: F00224
; CURRENT APPLICATION NUMBER: US/10/232,175
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US/09/710,249
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 501
; TYPE: PRT
; ORGANISM: human
US-10-232-175-17

Query Match      97.0%; Score 1149; DB 14; Length 501;
Best Local Similarity 93.6%; Pred. No. 1.4e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY      1  GPP-----GEPGTGLPGPPGGRGSGRFPDAGDVAGPKGPGAGRGSPGPA 48
Db      280  GPPGAGRGKRGARGEPGTGLPGPPGGRGSGRFPDAGDVAGPKGPGAGRGSPGPA 339
QY      49  GPKGSGPGRGPRGAGLCAKLTGSPGSPGDPGKTGPPGAGQDGRPGPPGARGQA 108
Db      340  GPKGSGPGRGPRGAGLCAKLTGSPGSPGDPGKTGPPGAGQDGRPGPPGARGQA 399
QY      109  GVMGFPKGAAGEPKAGRGVPPGAVGAGKDGAGAGGPGGPGAGRGEGQGPAG 168
Db      400  GVMGFPKGAAGEPKAGRGVPPGAVGAGKDGAGAGGPGGPGAGRGEGQGPAG 459
QY      169  GSPGQGLPGPAGPPGKAGRGVPPGDLGAPGSPGAG 208
Db      460  GSPGQGLPGPAGPPGKAGRGVPPGDLGAPGSPGAG 499

```

```

; LENGTH: 821
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hu-4
US-10-658-989A-3

Query Match      99.5%; Score 1178; DB 17; Length 821;
Best Local Similarity 100.0%; Pred. No. 4.9e-58;
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GPPGEPGTGLPGPPGGRGSGRFPDAGDVAGPKGPGAGRGSPGAGPKGSPGAGRP 60
Db      1  GPPGEPGTGLPGPPGGRGSGRFPDAGDVAGPKGPGAGRGSPGAGPKGSPGAGRP 60
QY      61  GEAGLPKAGKLTGSPGSGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGA 120
Db      61  GEAGLPKAGKLTGSPGSGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGA 120
QY      121  GEPKAGRGVPPGPGAVGAGKDGAGAGGPGGPGAGRGEGQGPAGSPGQGLPGPA 180
Db      121  GEPKAGRGVPPGPGAVGAGKDGAGAGGPGGPGAGRGEGQGPAGSPGQGLPGPA 180
QY      181  GPPGAGKPGQGVPPGDLGAPGSPGAG 208
Db      181  GPPGAGKPGQGVPPGDLGAPGSPGAG 208

RESULT 4
US-10-232-175-17
; Sequence 17, Application US/10232175
; Publication No. US20030064074A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Kivirikko, Kari I.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Olsen, David R.
; APPLICANT: Polarek, James W.
; TITLE OF INVENTION: RECOMBINANT GELATINS IN VACCINES
; FILE REFERENCE: F00224
; CURRENT APPLICATION NUMBER: US/10/232,175
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US/09/710,249
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 501
; TYPE: PRT
; ORGANISM: human
US-10-232-175-17

Query Match      97.0%; Score 1149; DB 14; Length 501;
Best Local Similarity 93.6%; Pred. No. 1.4e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY      1  GPP-----GEPGTGLPGPPGGRGSGRFPDAGDVAGPKGPGAGRGSPGPA 48
Db      280  GPPGAGRGKRGARGEPGTGLPGPPGGRGSGRFPDAGDVAGPKGPGAGRGSPGPA 339
QY      49  GPKGSGPGRGPRGAGLCAKLTGSPGSPGDPGKTGPPGAGQDGRPGPPGARGQA 108
Db      340  GPKGSGPGRGPRGAGLCAKLTGSPGSPGDPGKTGPPGAGQDGRPGPPGARGQA 399
QY      109  GVMGFPKGAAGEPKAGRGVPPGAVGAGKDGAGAGGPGGPGAGRGEGQGPAG 168
Db      400  GVMGFPKGAAGEPKAGRGVPPGAVGAGKDGAGAGGPGGPGAGRGEGQGPAG 459
QY      169  GSPGQGLPGPAGPPGKAGRGVPPGDLGAPGSPGAG 208
Db      460  GSPGQGLPGPAGPPGKAGRGVPPGDLGAPGSPGAG 499

```


;; PRIOR FILING DATE: 2003-08-01
;; NUMBER OF SEQ ID NOS: 13
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 10
;; LENGTH: 1014
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-901-816A-10

Query Match 97.0%; Score 1149; DB 17; Length 1014;
Best Local Similarity 93.6%; Pred. No. 2.3e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGGRGSGRGGFGADGVAGPKGPPGARGSGPGPA 48
|||
Db 280 GPPGAGEGKRGARGEPGPTGLPGPPGGRGSGRGGFGADGVAGPKGPPGARGSGPGPA 339
|||
QY 49 GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPPGAGDGRPPGPPGARGQA 108
|||
Db 340 GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPPGAGDGRPPGPPGARGQA 399
|||
QY 109 GVMGFPKGKGAAGFPKGAGRGVFPFGAVGPKGDEAGAGQPPGPPGARGSGPGPA 168
|||
Db 400 GVMGFPKGKGAAGFPKGAGRGVFPFGAVGPKGDEAGAGQPPGPPGARGSGPGPA 459
|||
QY 169 GSPGFGQLPGPPGAGKPGEGQGVPGDGLGAPGSPGAG 208
|||
Db 460 GSPGFGQLPGPPGAGKPGEGQGVPGDGLGAPGSPGARG 499
|||

RESULT 9

US-10-901-816A-11
;; Sequence 11, Application US/10901816A
;; Publication No. US20050058703A1
;; GENERAL INFORMATION:
;; APPLICANT: Chang, Robert C.
;; APPLICANT: Olsen, David R.
;; APPLICANT: James, Polarek W.
;; APPLICANT: Williams, Kim E.
;; TITLE OF INVENTION: Gelatin Capsules
;; FILE REFERENCE: FP0404 US
;; CURRENT APPLICATION NUMBER: US/10/901,816A
;; PRIOR FILING DATE: 2004-07-29
;; PRIOR APPLICATION NUMBER: US 60/492,085
;; PRIOR FILING DATE: 2003-08-01
;; NUMBER OF SEQ ID NOS: 13
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 11
;; LENGTH: 1014
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-901-816A-11

Query Match 97.0%; Score 1149; DB 17; Length 1014;
Best Local Similarity 93.6%; Pred. No. 2.3e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGGRGSGRGGFGADGVAGPKGPPGARGSGPGPA 48
|||
Db 280 GPPGAGEGKRGARGEPGPTGLPGPPGGRGSGRGGFGADGVAGPKGPPGARGSGPGPA 339
|||
QY 49 GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPPGAGDGRPPGPPGARGQA 108
|||
Db 340 GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPPGAGDGRPPGPPGARGQA 399
|||
QY 109 GVMGFPKGKGAAGFPKGAGRGVFPFGAVGPKGDEAGAGQPPGPPGARGSGPGPA 168
|||
Db 400 GVMGFPKGKGAAGFPKGAGRGVFPFGAVGPKGDEAGAGQPPGPPGARGSGPGPA 459
|||
QY 169 GSPGFGQLPGPPGAGKPGEGQGVPGDGLGAPGSPGAG 208
|||
Db 460 GSPGFGQLPGPPGAGKPGEGQGVPGDGLGAPGSPGARG 499
|||

RESULT 10

US-10-104-889-16
;; Sequence 16, Application US/10104889
;; Publication No. US20040086961A1
;; GENERAL INFORMATION:
;; APPLICANT: GRUSKIN, ELLIOT A.
;; BUECHTER, DOUGLAS
;; BROKAW, JANE
;; ZHANG, GUANGHUI
;; PAOLELLA, DAVID

TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE
STATE: NY
COUNTRY: U.S.A.
ZIP: 11553

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/104,889
FILING DATE: 22-Mar-2002

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998

ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516

INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1057 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown

MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-10-104-889-16

Query Match 97.0%; Score 1149; DB 15; Length 1057;
Best Local Similarity 93.6%; Pred. No. 2.4e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGPTGLPGPPGGRGSGRGGFGADGVAGPKGPPGARGSGPGPA 48
|||
Db 297 GPPGAGEGKRGARGEPGPTGLPGPPGGRGSGRGGFGADGVAGPKGPPGARGSGPGPA 356
|||
QY 49 GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPPGAGDGRPPGPPGARGQA 108
|||
Db 357 GPKSGPGEAGRPGEAGLPGAKGLTSGPSGPDGKTGPPGAGDGRPPGPPGARGQA 416
|||
QY 109 GVMGFPKGKGAAGFPKGAGRGVFPFGAVGPKGDEAGAGQPPGPPGARGSGPGPA 168
|||
Db 417 GVMGFPKGKGAAGFPKGAGRGVFPFGAVGPKGDEAGAGQPPGPPGARGSGPGPA 476
|||
QY 169 GSPGFGQLPGPPGAGKPGEGQGVPGDGLGAPGSPGAG 208
|||
Db 477 GSPGFGQLPGPPGAGKPGEGQGVPGDGLGAPGSPGARG 516
|||

RESULT 11

US-10-104-889-20
;; Sequence 20, Application US/10104889
;; Publication No. US20040086961A1
;; GENERAL INFORMATION:

```

; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
;
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
;
; NUMBER OF SEQUENCES: 50
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
;
; INFORMATION FOR SEQ ID NO: 20:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
;
; US-10-104-889-20
;
; Query Match 97.0%; Score 1149; DB 15; Length 1057;
; Best Local Similarity 93.6%; Pred. No. 2.4e-56;
; Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
;
; Qy 1 GPP-----GPPGPTGLPGERGSGRFPFGADGVAGPKGPGAGSGSPGA 48
; Db 297 GPPGAGEGKRGARGEPGPTGLPGERGSGRFPFGADGVAGPKGPGAGSGSPGA 356
;
; Qy 49 GPKGSPGEAGRPGEAGLPGAKGLTSGSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
; Db 357 GPKGSPGEAGRPGEAGLPGAKGLTSGSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 416
;
; Qy 109 GVMGFPKGAAGRPKAGRGVPPGAVGPKAGKDGAGAGQGGPPGAGPAGERGEOGPA 168
; Db 417 GVMGFPKGAAGRPKAGRGVPPGAVGPKAGKDGAGAGQGGPPGAGPAGERGEOGPA 476
;
; Qy 169 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 208
; Db 477 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 516
;
; RESULT 12
; US-10-104-889-11
; Sequence 11, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
;
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
;
; NUMBER OF SEQUENCES: 50
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
;
; INFORMATION FOR SEQ ID NO: 20:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
;
; US-10-104-889-20
;
; Query Match 97.0%; Score 1149; DB 15; Length 1057;
; Best Local Similarity 93.6%; Pred. No. 2.4e-56;
; Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
;
; Qy 1 GPP-----GPPGPTGLPGERGSGRFPFGADGVAGPKGPGAGSGSPGA 48
; Db 297 GPPGAGEGKRGARGEPGPTGLPGERGSGRFPFGADGVAGPKGPGAGSGSPGA 356
;
; Qy 49 GPKGSPGEAGRPGEAGLPGAKGLTSGSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
; Db 357 GPKGSPGEAGRPGEAGLPGAKGLTSGSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 416
;
; Qy 109 GVMGFPKGAAGRPKAGRGVPPGAVGPKAGKDGAGAGQGGPPGAGPAGERGEOGPA 168
; Db 417 GVMGFPKGAAGRPKAGRGVPPGAVGPKAGKDGAGAGQGGPPGAGPAGERGEOGPA 476
;
; Qy 169 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 208
; Db 477 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 516
;
; RESULT 13
; US-10-104-889-6
; Sequence 6, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
;
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
;
; NUMBER OF SEQUENCES: 50
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
;
; INFORMATION FOR SEQ ID NO: 11:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1107 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
;
; US-10-104-889-11
;
; Query Match 97.0%; Score 1149; DB 15; Length 1107;
; Best Local Similarity 93.6%; Pred. No. 2.5e-56;
; Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
;
; Qy 1 GPP-----GPPGPTGLPGERGSGRFPFGADGVAGPKGPGAGSGSPGA 48
; Db 297 GPPGAGEGKRGARGEPGPTGLPGERGSGRFPFGADGVAGPKGPGAGSGSPGA 356
;
; Qy 49 GPKGSPGEAGRPGEAGLPGAKGLTSGSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
; Db 357 GPKGSPGEAGRPGEAGLPGAKGLTSGSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 416
;
; Qy 109 GVMGFPKGAAGRPKAGRGVPPGAVGPKAGKDGAGAGQGGPPGAGPAGERGEOGPA 168
; Db 417 GVMGFPKGAAGRPKAGRGVPPGAVGPKAGKDGAGAGQGGPPGAGPAGERGEOGPA 476
;
; Qy 169 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 208
; Db 477 GSPGFQGLPGPAGPPGAGKPGEGQVPGDLGAPGSPGAG 516
;
; RESULT 13
; US-10-104-889-6
; Sequence 6, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
;
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
;
; NUMBER OF SEQUENCES: 50
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
;
; INFORMATION FOR SEQ ID NO: 11:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1107 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
;
; US-10-104-889-11

```

```
/
/ COUNTRY: U.S.A.
/ ZIP: 11553
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/104,889
/ FILING DATE: 22-Mar-2002
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/09/169,768
/ FILING DATE: 09-OCT-1998
/ ATTORNEY/AGENT INFORMATION:
/ NAME: STEEN, JEFFREY S
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (516) 228-8484
/ TELEFAX: (516) 228-8516
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1169 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: unknown
/ MOLECULE TYPE: peptide
/ SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-104-889-6

Query Match 97.0%; Score 1149; DB 15; Length 1169;
Best Local Similarity 93.6%; Pred. No. 2.6e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGTLGPPGGRGFGFPGADGVAGPKPAGERGSPGPA 48
Db 297 GPPGAGEGKRGARGEPGTLGPPGGRGFGFPGADGVAGPKPAGERGSPGPA 356
QY 49 GPKSGPEAGRPEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
Db 357 GPKSGPEAGRPEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 416
QY 109 GWMGFPKGAAGEPKGAGRGVPPGCAVGPAGKDGAGACQPPGAPGAGERGEOGPA 168
Db 417 GWMGFPKGAAGEPKGAGRGVPPGCAVGPAGKDGAGACQPPGAPGAGERGEOGPA 476
QY 169 GSPGFQGLPGPAGPCEAGKPGEGQGVPGDLGAPGSGPAG 208
Db 477 GSPGFQGLPGPAGPCEAGKPGEGQGVPGDLGAPGSGPAG 516

RESULT 14
US-10-104-889-8
/ Sequence 8, Application US/10104889
/ Publication No. US20040086961A1
/ GENERAL INFORMATION:
/ APPLICANT: GRUSKIN, ELLIOT A.
/ BUECHTER, DOUGLAS
/ BROKAW, JANE
/ ZHANG, GUANGHUI
/ PAOLELLA, DAVID
/ TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
/ NUMBER OF SEQUENCES: 50
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: DILNORTH & BARRESE
/ STREET: 333 EARLE OVINGTON BOULEVARD
/ CITY: UNIONDALE
/ STATE: NY
/ COUNTRY: U.S.A.
/ ZIP: 11553
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
/
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/104,889
/ FILING DATE: 22-Mar-2002
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/09/169,768
/ FILING DATE: 09-OCT-1998
/ ATTORNEY/AGENT INFORMATION:
/ NAME: STEEN, JEFFREY S
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (516) 228-8484
/ TELEFAX: (516) 228-8516
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1171 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: unknown
/ MOLECULE TYPE: peptide
/ SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-104-889-8

Query Match 97.0%; Score 1149; DB 15; Length 1171;
Best Local Similarity 93.6%; Pred. No. 2.6e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPGTLGPPGGRGFGFPGADGVAGPKPAGERGSPGPA 48
Db 297 GPPGAGEGKRGARGEPGTLGPPGGRGFGFPGADGVAGPKPAGERGSPGPA 356
QY 49 GPKSGPEAGRPEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
Db 357 GPKSGPEAGRPEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 416
QY 109 GWMGFPKGAAGEPKGAGRGVPPGCAVGPAGKDGAGACQPPGAPGAGERGEOGPA 168
Db 417 GWMGFPKGAAGEPKGAGRGVPPGCAVGPAGKDGAGACQPPGAPGAGERGEOGPA 476
QY 169 GSPGFQGLPGPAGPCEAGKPGEGQGVPGDLGAPGSGPAG 208
Db 477 GSPGFQGLPGPAGPCEAGKPGEGQGVPGDLGAPGSGPAG 516

RESULT 15
US-10-058-124-18
/ Sequence 18, Application US/10058124
/ Publication No. US20030119058A1
/ GENERAL INFORMATION:
/ APPLICANT: Qvist, Per
/ Bonde, Martin
/ TITLE OF INVENTION: A Method for Assaying Collagen Fragments
/ in Body Fluids, A Test Kit and Means for Carrying Out the
/ Method and Use of the Method to Diagnose the Presence of
/ Disorders Associated with the Metabolism of
/ NUMBER OF SEQUENCES: 21
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Darby & Darby PC
/ STREET: 805 Third Avenue
/ CITY: New York
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10022
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/058,124
/ FILING DATE: 29-Jan-2002
/ CLASSIFICATION: <Unknown>
/ PRIOR APPLICATION DATA:
```

```

; APPLICATION NUMBER: 09/570,573
; FILING DATE: 2002-MAY-12
; APPLICATION NUMBER: 08/187,319
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
;
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-058-124-18

```

```

Query Match      97.0%; Score 1149; DB 14; Length 1341;
Best Local Similarity 93.6%; Pred. No. 2.9e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GERPPTGLPGPGGSGRGGPGADGVAGPKGPGAGRGSPGPA 48
Db |||
334 GPPGAGEGKRGAGEPGPTGLPGPGGSGRGGPGADGVAGPKGPGAGRGSPGPA 393
QY 49 GPKGSPGEAGRPGEAGLPGAKGLTSGPSGPGDGKTGPPGAGQDGRPPGPPGARGQA 108
Db |||
394 GPKGSPGEAGRPGEAGLPGAKGLTSGPSGPGDGKTGPPGAGQDGRPPGPPGARGQA 453
QY 109 GVMGFPKGAAGBPGKAGRGVPGPGAVGPAKDGAGAGCGPPGAGPAGERGEGQGPA 168
Db |||
454 GVMGFPKGAAGBPGKAGRGVPGPGAVGPAKDGAGAGCGPPGAGPAGERGEGQGPA 513
QY 169 GSRPGFGLPGAGPGEAGKPGEGVPGDLGAPGSPGAG 208
Db |||
514 GSPGFQGLPGAGPGEAGKPGEGVPGDLGAPGSPGAG 553

```

Search completed: June 17, 2005, 15:35:29
Job time : 24.8937 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:42:55 ; Search time 7.15427 Seconds
(without alignments)
2810.812 Million cell updates/sec

Title: US-10-658-989A-1
Perfect score: 1184
Sequence: 1 GPPGEPGTGLPGPPGSGG.....GEQGVGDLGNPGSPGAGG 209

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1149	97.0	1464	1 CGHUIS	collagen alpha 1(I)
2	1118	94.4	671	1 CGRTIS	collagen alpha 1(I)
3	1116	94.3	1453	2 S21626	collagen alpha 1(I)
4	1067	90.1	1042	1 CGHUIS	collagen alpha 1(I)
5	862	72.8	1487	1 CGHU6C	collagen alpha 1(I)
6	860	72.6	1418	2 T45467	collagen alpha 1(I)
7	850	71.8	1486	1 B40333	collagen alpha 1(I)
8	847	71.5	1492	2 A40333	collagen alpha 1(I)
9	839	70.9	1419	2 A41182	collagen alpha 1(I)
10	839	70.9	1487	2 B41182	collagen alpha 1(I)
11	824	69.6	779	1 CGB01S	collagen alpha 1(I)
12	815	68.8	673	1 CGB06C	collagen alpha 1(I)
13	794.5	67.1	886	2 I50694	collagen alpha 1(I)
14	786	66.4	1496	1 CGHU2V	collagen alpha 2(I)
15	776	65.5	1466	1 CGHU7L	collagen alpha 1(I)
16	765	64.6	1497	2 I49607	procollagen type V
17	763	64.4	1464	2 S59856	collagen alpha 1(I)
18	761.5	64.3	1049	1 CGB07S	collagen alpha 1(I)
19	760	64.2	1373	1 A43291	collagen alpha 2(I)
20	744	62.8	1366	1 CGHU2S	collagen alpha 2(I)
21	629.5	53.2	1027	2 S28774	collagen alpha cha
22	607	51.3	1838	1 CGHU1V	collagen alpha 1(I)
23	602.5	50.9	310	2 I50696	collagen alpha 1(I)
24	600.5	50.7	1414	1 S23809	collagen alpha 2(I)
25	599	50.6	1843	2 S18803	collagen alpha 1(I)
26	590.5	49.9	632	2 S42731	collagen alpha 1 c
27	589	49.7	964	1 CGCH2S	collagen alpha 2(I)
28	577	48.7	1806	1 CGHU1E	collagen alpha 1(X)
29	576.5	48.7	730	2 A36226	collagen alpha 1 c

collagen alpha 5(I)
collagen alpha 1(X)
collagen alpha 1(X)
collagen alpha 3(I)
collagen alpha 4(I)
collagen alpha 1(I)
collagen alpha 2(X)
collagen alpha 1(I)
collagen alpha 1(V)
collagen alpha 5(I)
collagen alpha 1(I)
collagen alpha 1(X)
type VII collagen
collagen alpha 1(I)

30 576 48.6 754 2 A55267
31 576 48.6 1024 2 S18251
32 575.5 48.6 3124 2 A40020
33 575 48.6 888 2 S28791
34 573.5 48.4 675 2 S20819
35 573.5 48.4 1690 1 CGHU1B
36 572.5 48.4 636 2 S41067
37 572.5 48.4 1546 1 CGHU2E
38 572 48.3 615 2 A05269
39 563.5 47.6 2944 2 A54849
40 563 47.6 1691 1 S22917
41 562.5 47.5 488 2 A27353
42 561.5 47.4 674 2 S23297
43 561 47.4 1603 2 S23810
44 557.5 47.1 1549 2 I48103
45 555.5 46.9 931 2 S13580

ALIGNMENTS

RESULT 1
CGHUIS
collagen alpha 1(I) chain precursor - human
N:Alternate names: procollagen alpha 1(I) chain
C:Species: Homo sapiens (man)
C:Date: 12-Aug-1981 #sequence_revision 04-Oct-1996 #text_change 09-Jul-2004
C:Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S11:
5269; A29439; I53466; A02853; I37247
R:D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.
Gene 67, 105-115, 1988
A:Title: Complete nucleotide sequence of the region encompassing the first twenty-five ex
A:Reference number: I60114; MUID:88329734; PMID:2843432
A:Accession: I60114
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-369, 'L', 371-589 <DAL>
A:Cross-references: UNIPROT:P02452; UNIPROT:Q14992; UNIPROT:Q16053; UNIPROT:Q13896; UNIP:
R:Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Juenisch, R.; Procke
Biochem. J. 253, 919-922, 1988
A:Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human t
A:Reference number: S01143; MUID:89025644; PMID:3178743
A:Accession: S01143
A:Molecule type: mRNA
A:Residues: 1-472 <TRO>
A:Cross-references: EMBL:X07884; NID:g30015; PIDN:CAA30731.1; PID:g30016; GB:W36546; NID:
A:Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.;
Nature 310, 337-340, 1984
A:Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of
A:Reference number: A93335; MUID:84270697; PMID:6462220
A:Accession: A93335
A:Molecule type: DNA
A:Residues: 1-58, 'Q', 60-181 <CHU>
A:Cross-references: EMBL:X00820; NID:g35657; PIDN:CAA25394.1; PID:g35658
R:Rossouw, C.M.S.; Vergeer, W.P.; du Flooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.;
J. Biol. Chem. 262, 15151-15157, 1987
A:Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enha
A:Reference number: I55254; MUID:88033098; PMID:2822714
A:Accession: I55254
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-45 <ROS>
A:Cross-references: GB:J02829; NID:g180387; PIDN:AAA51993.1; PID:g180388
R:Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A:Title: Regulatory elements in the first intron contribute to transcriptional control of
A:Reference number: A39943; MUID:88097389; PMID:3480516
A:Accession: A39943
A:Molecule type: DNA
A:Residues: 1-34 <BOR>
A:Cross-references: GB:J03559; NID:g180876; PIDN:AAA52052.1; PID:g553238
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

J. Biol. Chem. 260, 2315-2320, 1985
A/Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter structure and organization of the 5' and 3' ends only are shown; mutant sequence 263-Arg reported in the literature.
A/Reference number: I55237; MUID:05130970; PMID:2857713
A/Accession: I55237
A/Status: translation not shown; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-34 <CH2>
A/Cross-references: GB:M10627; NID:g180383; PIDN:AA51992.1; PID:g553226
R/Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollister, J. Biol. Chem. 265, 6312-6317, 1990
A/Title: In vivo and in vitro noncovalent association of excised alpha 1(I) amino-terminal propeptides of type I collagen.
A/Reference number: A35233; MUID:90202908; PMID:2318855
A/Accession: A35233
A/Molecule type: protein
A/Residues: 33-52 <WTR>
A/Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved propeptide.
R/Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F. EMBO J. 8, 1705-1710, 1989
A/Title: A base substitution in the exon of a collagen gene causes alternative splicing of the alpha 1(I) chain of type I collagen.
A/Reference number: S09400; MUID:89356643; PMID:2767050
A/Accession: S09400
A/Molecule type: mRNA
A/Residues: 156-183 <WET>
R/Click, E.M.; Bornstein, P. Biochemistry 9, 4699-4706, 1970
A/Title: Isolation and characterization of the cyanogen bromide peptides from the alpha 1(I) chain of type I collagen.
A/Reference number: A90567; MUID:71038625; PMID:5529814
A/Contents: CNBR0-1, CNBR2, CNBR4, CNBR5
A/Accession: B90567
A/Molecule type: protein
A/Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z', 235, 'Z', 238, 'Z', 240-241, 'Z', 243-244, 'Z', 246-247, 'Z', 249-250, 'Z', 252-253, 'Z', 255-256, 'Z', 258-259, 'Z', 261-262, 'Z', 264-265, 'Z', 267-268, 'Z', 270-271, 'Z', 273-274, 'Z', 276-277, 'Z', 279-280, 'Z', 282-283, 'Z', 285-286, 'Z', 288-289, 'Z', 291-292, 'Z', 294-295, 'Z', 297-298, 'Z', 300-301, 'Z', 303-304, 'Z', 306-307, 'Z', 309-310, 'Z', 312-313, 'Z', 315-316, 'Z', 318-319, 'Z', 321-322, 'Z', 324-325, 'Z', 327-328, 'Z', 330-331, 'Z', 333-334, 'Z', 336-337, 'Z', 339-340, 'Z', 342-343, 'Z', 345-346, 'Z', 348-349, 'Z', 351-352, 'Z', 354-355, 'Z', 357-358, 'Z', 360-361, 'Z', 363-364, 'Z', 366-367, 'Z', 369-370, 'Z', 372-373, 'Z', 375-376, 'Z', 378-379, 'Z', 381-382, 'Z', 384-385, 'Z', 387-388, 'Z', 390-391, 'Z', 393-394, 'Z', 396-397, 'Z', 399-400, 'Z', 402-403, 'Z', 405-406, 'Z', 408-409, 'Z', 411-412, 'Z', 414-415, 'Z', 417-418, 'Z', 420-421, 'Z', 423-424, 'Z', 426-427, 'Z', 429-430, 'Z', 432-433, 'Z', 435-436, 'Z', 438-439, 'Z', 441-442, 'Z', 444-445, 'Z', 447-448, 'Z', 450-451, 'Z', 453-454, 'Z', 456-457, 'Z', 459-460, 'Z', 462-463, 'Z', 465-466, 'Z', 468-469, 'Z', 471-472, 'Z', 474-475, 'Z', 477-478, 'Z', 480-481, 'Z', 483-484, 'Z', 486-487, 'Z', 489-490, 'Z', 492-493, 'Z', 495-496, 'Z', 498-499, 'Z', 501-502, 'Z', 504-505, 'Z', 507-508, 'Z', 510-511, 'Z', 513-514, 'Z', 516-517, 'Z', 519-520, 'Z', 522-523, 'Z', 524-525, 'Z', 527-528, 'Z', 530-531, 'Z', 533-534, 'Z', 536-537, 'Z', 539-540, 'Z', 542-543, 'Z', 545-546, 'Z', 548-549, 'Z', 551-552, 'Z', 554-555, 'Z', 557-558, 'Z', 560-561, 'Z', 563-564, 'Z', 566-567, 'Z', 569-570, 'Z', 572-573, 'Z', 575-576, 'Z', 578-579, 'Z', 581-582, 'Z', 584-585, 'Z', 587-588, 'Z', 590-591, 'Z', 593-594, 'Z', 596-597, 'Z', 599-600, 'Z', 602-603, 'Z', 605-606, 'Z', 608-609, 'Z', 611-612, 'Z', 614-615, 'Z', 617-618, 'Z', 620-621, 'Z', 623-624, 'Z', 626-627, 'Z', 629-630, 'Z', 632-633, 'Z', 635-636, 'Z', 638-639, 'Z', 641-642, 'Z', 644-645, 'Z', 647-648, 'Z', 650-651, 'Z', 653-654, 'Z', 656-657, 'Z', 659-660, 'Z', 662-663, 'Z', 665-666, 'Z', 668-669, 'Z', 671-672, 'Z', 674-675, 'Z', 677-678, 'Z', 680-681, 'Z', 683-684, 'Z', 686-687, 'Z', 689-690, 'Z', 692-693, 'Z', 695-696, 'Z', 698-699, 'Z', 701-702, 'Z', 704-705, 'Z', 707-708, 'Z', 710-711, 'Z', 713-714, 'Z', 716-717, 'Z', 719-720, 'Z', 722-723, 'Z', 725-726, 'Z', 728-729, 'Z', 731-732, 'Z', 734-735, 'Z', 737-738, 'Z', 740-741, 'Z', 743-744, 'Z', 746-747, 'Z', 749-750, 'Z', 752-753, 'Z', 755-756, 'Z', 758-759, 'Z', 761-762, 'Z', 764-765, 'Z', 767-768, 'Z', 770-771, 'Z', 773-774, 'Z', 776-777, 'Z', 779-780, 'Z', 782-783, 'Z', 785-786, 'Z', 788-789, 'Z', 791-792, 'Z', 794-795, 'Z', 797-798, 'Z', 800-801, 'Z', 803-804, 'Z', 806-807, 'Z', 809-810, 'Z', 812-813, 'Z', 815-816, 'Z', 818-819, 'Z', 821-822, 'Z', 824-825, 'Z', 827-828, 'Z', 830-831, 'Z', 833-834, 'Z', 836-837, 'Z', 839-840, 'Z', 842-843, 'Z', 845-846, 'Z', 848-849, 'Z', 851-852, 'Z', 854-855, 'Z', 857-858, 'Z', 860-861, 'Z', 863-864, 'Z', 866-867, 'Z', 869-870, 'Z', 872-873, 'Z', 875-876, 'Z', 878-879, 'Z', 881-882, 'Z', 884-885, 'Z', 887-888, 'Z', 890-891, 'Z', 893-894, 'Z', 896-897, 'Z', 899-900, 'Z', 902-903, 'Z', 905-906, 'Z', 908-909, 'Z', 911-912, 'Z', 914-915, 'Z', 917-918, 'Z', 920-921, 'Z', 923-924, 'Z', 926-927, 'Z', 929-930, 'Z', 932-933, 'Z', 935-936, 'Z', 938-939, 'Z', 941-942, 'Z', 944-945, 'Z', 947-948, 'Z', 950-951, 'Z', 953-954, 'Z', 956-957, 'Z', 959-960, 'Z', 962-963, 'Z', 965-966, 'Z', 968-969, 'Z', 971-972, 'Z', 974-975, 'Z', 977-978, 'Z', 980-981, 'Z', 983-984, 'Z', 986-987, 'Z', 989-990, 'Z', 992-993, 'Z', 995-996, 'Z', 998-999, 'Z', 1000-1001, 'Z', 1003-1004, 'Z', 1006-1007, 'Z', 1009-1010, 'Z', 1012-1013, 'Z', 1015-1016, 'Z', 1018-1019, 'Z', 1021-1022, 'Z', 1024-1025, 'Z', 1027-1028, 'Z', 1030-1031, 'Z', 1033-1034, 'Z', 1036-1037, 'Z', 1039-1040, 'Z', 1042-1043, 'Z', 1045-1046, 'Z', 1048-1049, 'Z', 1051-1052, 'Z', 1054-1055, 'Z', 1057-1058, 'Z', 1060-1061, 'Z', 1063-1064, 'Z', 1066-1067, 'Z', 1069-1070, 'Z', 1072-1073, 'Z', 1075-1076, 'Z', 1078-1079, 'Z', 1081-1082, 'Z', 1084-1085, 'Z', 1087-1088, 'Z', 1090-1091, 'Z', 1093-1094, 'Z', 1096-1097, 'Z', 1099-1100, 'Z', 1102-1103, 'Z', 1105-1106, 'Z', 1108-1109, 'Z', 1111-1112, 'Z', 1114-1115, 'Z', 1117-1118, 'Z', 1120-1121, 'Z', 1123-1124, 'Z', 1126-1127, 'Z', 1129-1130, 'Z', 1132-1133, 'Z', 1135-1136, 'Z', 1138-1139, 'Z', 1141-1142, 'Z', 1144-1145, 'Z', 1147-1148, 'Z', 1150-1151, 'Z', 1153-1154, 'Z', 1156-1157, 'Z', 1159-1160, 'Z', 1162-1163, 'Z', 1165-1166, 'Z', 1168-1169, 'Z', 1171-1172, 'Z', 1174-1175, 'Z', 1177-1178, 'Z', 1180-1181, 'Z', 1183-1184, 'Z', 1186-1187, 'Z', 1189-1190, 'Z', 1192-1193, 'Z', 1195-1196, 'Z', 1198-1199, 'Z', 1201-1202, 'Z', 1204-1205, 'Z', 1207-1208, 'Z', 1210-1211, 'Z', 1213-1214, 'Z', 1216-1217, 'Z', 1219-1220, 'Z', 1222-1223, 'Z', 1225-1226, 'Z', 1228-1229, 'Z', 1231-1232, 'Z', 1234-1235, 'Z', 1237-1238, 'Z', 1240-1241, 'Z', 1243-1244, 'Z', 1246-1247, 'Z', 1249-1250, 'Z', 1252-1253, 'Z', 1255-1256, 'Z', 1258-1259, 'Z', 1261-1262, 'Z', 1264-1265, 'Z', 1267-1268, 'Z', 1270-1271, 'Z', 1273-1274, 'Z', 1276-1277, 'Z', 1279-1280, 'Z', 1282-1283, 'Z', 1285-1286, 'Z', 1288-1289, 'Z', 1291-1292, 'Z', 1294-1295, 'Z', 1297-1298, 'Z', 1300-1301, 'Z', 1303-1304, 'Z', 1306-1307, 'Z', 1309-1310, 'Z', 1312-1313, 'Z', 1315-1316, 'Z', 1318-1319, 'Z', 1321-1322, 'Z', 1324-1325, 'Z', 1327-1328, 'Z', 1330-1331, 'Z', 1333-1334, 'Z', 1336-1337, 'Z', 1339-1340, 'Z', 1342-1343, 'Z', 1345-1346, 'Z', 1348-1349, 'Z', 1351-1352, 'Z', 1354-1355, 'Z', 1357-1358, 'Z', 1360-1361, 'Z', 1363-1364, 'Z', 1366-1367, 'Z', 1369-1370, 'Z', 1372-1373, 'Z', 1375-1376, 'Z', 1378-1379, 'Z', 1381-1382, 'Z', 1384-1385, 'Z', 1387-1388, 'Z', 1390-1391, 'Z', 1393-1394, 'Z', 1396-1397, 'Z', 1399-1400, 'Z', 1402-1403, 'Z', 1405-1406, 'Z', 1408-1409, 'Z', 1411-1412, 'Z', 1414-1415, 'Z', 1417-1418, 'Z', 1420-1421, 'Z', 1423-1424, 'Z', 1426-1427, 'Z', 1429-1430, 'Z', 1432-1433, 'Z', 1435-1436, 'Z', 1438-1439, 'Z', 1441-1442, 'Z', 1444-1445, 'Z', 1447-1448, 'Z', 1450-1451, 'Z', 1453-1454, 'Z', 1456-1457, 'Z', 1459-1460, 'Z', 1462-1463, 'Z', 1465-1466, 'Z', 1468-1469, 'Z', 1471-1472, 'Z', 1474-1475, 'Z', 1477-1478, 'Z', 1480-1481, 'Z', 1483-1484, 'Z', 1486-1487, 'Z', 1489-1490, 'Z', 1492-1493, 'Z', 1495-1496, 'Z', 1498-1499, 'Z', 1501-1502, 'Z', 1504-1505, 'Z', 1507-1508, 'Z', 1510-1511, 'Z', 1513-1514, 'Z', 1516-1517, 'Z', 1519-1520, 'Z', 1522-1523, 'Z', 1524-1525, 'Z', 1527-1528, 'Z', 1530-1531, 'Z', 1533-1534, 'Z', 1536-1537, 'Z', 1539-1540, 'Z', 1542-1543, 'Z', 1545-1546, 'Z', 1548-1549, 'Z', 1551-1552, 'Z', 1554-1555, 'Z', 1557-1558, 'Z', 1560-1561, 'Z', 1563-1564, 'Z', 1566-1567, 'Z', 1569-1570, 'Z', 1572-1573, 'Z', 1575-1576, 'Z', 1578-1579, 'Z', 1581-1582, 'Z', 1584-1585, 'Z', 1587-1588, 'Z', 1590-1591, 'Z', 1593-1594, 'Z', 1596-1597, 'Z', 1599-1600, 'Z', 1602-1603, 'Z', 1605-1606, 'Z', 1608-1609, 'Z', 1611-1612, 'Z', 1614-1615, 'Z', 1617-1618, 'Z', 1620-1621, 'Z', 1623-1624, 'Z', 1626-1627, 'Z', 1629-1630, 'Z', 1632-1633, 'Z', 1635-1636, 'Z', 1638-1639, 'Z', 1641-1642, 'Z', 1644-1645, 'Z', 1647-1648, 'Z', 1650-1651, 'Z', 1653-1654, 'Z', 1656-1657, 'Z', 1659-1660, 'Z', 1662-1663, 'Z', 1665-1666, 'Z', 1668-1669, 'Z', 1671-1672, 'Z', 1674-1675, 'Z', 1677-1678, 'Z', 1680-1681, 'Z', 1683-1684, 'Z', 1686-1687, 'Z', 1689-1690, 'Z', 1692-1693, 'Z', 1695-1696, 'Z', 1698-1699, 'Z', 1701-1702, 'Z', 1704-1705, 'Z', 1707-1708, 'Z', 1710-1711, 'Z', 1713-1714, 'Z', 1716-1717, 'Z', 1719-1720, 'Z', 1722-1723, 'Z', 1725-1726, 'Z', 1728-1729, 'Z', 1731-1732, 'Z', 1734-1735, 'Z', 1737-1738, 'Z', 1740-1741, 'Z', 1743-1744, 'Z', 1746-1747, 'Z', 1749-1750, 'Z', 1752-1753, 'Z', 1755-1756, 'Z', 1758-1759, 'Z', 1761-1762, 'Z', 1764-1765, 'Z', 1767-1768, 'Z', 1770-1771, 'Z', 1773-1774, 'Z', 1776-1777, 'Z', 1779-1780, 'Z', 1782-1783, 'Z', 1785-1786, 'Z', 1788-1789, 'Z', 1791-1792, 'Z', 1794-1795, 'Z', 1797-1798, 'Z', 1800-1801, 'Z', 1803-1804, 'Z', 1806-1807, 'Z', 1809-1810, 'Z', 1812-1813, 'Z', 1815-1816, 'Z', 1818-1819, 'Z', 1821-1822, 'Z', 1824-1825, 'Z', 1827-1828, 'Z', 1830-1831, 'Z', 1833-1834, 'Z', 1836-1837, 'Z', 1839-1840, 'Z', 1842-1843, 'Z', 1845-1846, 'Z', 1848-1849, 'Z', 1851-1852, 'Z', 1854-1855, 'Z', 1857-1858, 'Z', 1860-1861, 'Z', 1863-1864, 'Z', 1866-1867, 'Z', 1869-1870, 'Z', 1872-1873, 'Z', 1875-1876, 'Z', 1878-1879, 'Z', 1881-1882, 'Z', 1884-1885, 'Z', 1887-1888, 'Z', 1890-1891, 'Z', 1893-1894, 'Z', 1896-1897, 'Z', 1899-1900, 'Z', 1902-1903, 'Z', 1905-1906, 'Z', 1908-1909, 'Z', 1911-1912, 'Z', 1914-1915, 'Z', 1917-1918, 'Z', 1920-1921, 'Z', 1923-1924, 'Z', 1926-1927, 'Z', 1929-1930, 'Z', 1932-1933, 'Z', 1935-1936, 'Z', 1938-1939, 'Z', 1941-1942, 'Z', 1944-1945, 'Z', 1947-1948, 'Z', 1950-1951, 'Z', 1953-1954, 'Z', 1956-1957, 'Z', 1959-1960, 'Z', 1962-1963, 'Z', 1965-1966, 'Z', 1968-1969, 'Z', 1971-1972, 'Z', 1974-1975, 'Z', 1977-1978, 'Z', 1980-1981, 'Z', 1983-1984, 'Z', 1986-1987, 'Z', 1989-1990, 'Z', 1992-1993, 'Z', 1995-1996, 'Z', 1998-1999, 'Z', 2000-2001, 'Z', 2003-2004, 'Z', 2006-2007, 'Z', 2009-2010, 'Z', 2012-2013, 'Z', 2015-2016, 'Z', 2018-2019, 'Z', 2021-2022, 'Z', 2024-2025, 'Z', 2027-2028, 'Z', 2030-2031, 'Z', 2033-2034, 'Z', 2036-2037, 'Z', 2039-2040, 'Z', 2042-2043, 'Z', 2045-2046, 'Z', 2048-2049, 'Z', 2051-2052, 'Z', 2054-2055, 'Z', 2057-2058, 'Z', 2060-2061, 'Z', 2063-2064, 'Z', 2066-2067, 'Z', 2069-2070, 'Z', 2072-2073, 'Z', 2075-2076, 'Z', 2078-2079, 'Z', 2081-2082, 'Z', 2084-2085, 'Z', 2087-2088, 'Z', 2090-2091, 'Z', 2093-2094, 'Z', 2096-2097, 'Z', 2099-2100, 'Z', 2102-2103, 'Z', 2105-2106, 'Z', 2108-2109, 'Z', 2111-2112, 'Z', 2114-2115, 'Z', 2117-2118, 'Z', 2120-2121, 'Z', 2123-2124, 'Z', 2126-2127, 'Z', 2129-2130, 'Z', 2132-2133, 'Z', 2135-2136, 'Z', 2138-2139, 'Z', 2141-2142, 'Z', 2144-2145, 'Z', 2147-2148, 'Z', 2150-2151, 'Z', 2153-2154, 'Z', 2156-2157, 'Z', 2159-2160, 'Z', 2162-2163, 'Z', 2165-2166, 'Z', 2168-2169, 'Z', 2171-2172, 'Z', 2174-2175, 'Z', 2177-2178, 'Z', 2180-2181, 'Z', 2183-2184, 'Z', 2186-2187, 'Z', 2189-2190, 'Z', 2192-2193, 'Z', 2195-2196, 'Z', 2198-2199, 'Z', 2201-2202, 'Z', 2204-2205, 'Z', 2207-2208, 'Z', 2210-2211, 'Z', 2213-2214, 'Z', 2216-2217, 'Z', 2219-2220, 'Z', 2222-2223, 'Z', 2225-2226, 'Z', 2228-2229, 'Z', 2231-2232, 'Z', 2234-2235, 'Z', 2237-2238, 'Z', 2240-2241, 'Z', 2243-2244, 'Z', 2246-2247, 'Z', 2249-2250, 'Z', 2252-2253, 'Z', 2255-2256, 'Z', 2258-2259, 'Z', 2261-2262, 'Z', 2264-2265, 'Z', 2267-2268, 'Z', 2270-2271, 'Z', 2273-2274, 'Z', 2276-2277, 'Z', 2279-2280, 'Z', 2282-2283, 'Z', 2285-2286, 'Z', 2288-2289, 'Z', 2291-2292, 'Z', 2294-2295, 'Z', 2297-2298, 'Z', 2300-2301, 'Z', 2303-2304, 'Z', 2306-2307, 'Z', 2309-2310, 'Z', 2312-2313, 'Z', 2315-2316, 'Z', 2318-2319, 'Z', 2321-2322, 'Z', 2324-2325, 'Z', 2327-2328, 'Z', 2330-2331, 'Z', 2333-2334, 'Z', 2336-2337, 'Z', 2339-2340, 'Z', 2342-2343, 'Z', 2345-2346, 'Z', 2348-2349, 'Z', 2351-2352, 'Z', 2354-2355, 'Z', 2357-2358, 'Z', 2360-2361, 'Z', 2363-2364, 'Z', 2366-2367, 'Z', 2369-2370, 'Z', 2372-2373, 'Z', 2375-2376, 'Z', 2378-2379, 'Z', 2381-2382, 'Z', 2384-2385, 'Z', 2387-2388, 'Z', 2390-2391, 'Z', 2393-2394, 'Z', 2396-2397, 'Z', 2399-2400, 'Z', 2402-2403, 'Z', 2405-2406, 'Z', 2408-2409, 'Z', 2411-2412, 'Z', 2414-2415, 'Z', 2417-2418, 'Z', 2420-2421, 'Z', 2423-2424, 'Z', 2426-2427, 'Z', 2429-2430, 'Z', 2432-2433, 'Z', 2435-2436, 'Z', 2438-2439, 'Z', 2441-2442, 'Z', 2444-2445, 'Z', 2447-2448, 'Z', 2450-2451, 'Z', 2453-2454, 'Z', 2456-2457, 'Z', 2459-2460, 'Z', 2462-2463, 'Z', 2465-2466, 'Z', 2468-2469, 'Z', 2471-2472, 'Z', 2474-2475, 'Z', 2477-2478, 'Z', 2480-2481, 'Z', 2483-2484, 'Z', 2486-2487, 'Z', 2489-2490, 'Z', 2492-2493, 'Z', 2495-2496, 'Z', 2498-2499, 'Z', 2501-2502, 'Z', 2504-2505, 'Z', 2507-2508, 'Z', 2510-2511, 'Z', 2513-2514, 'Z', 2516-2517, 'Z', 2519-2520, 'Z', 2522-2523, 'Z', 2524-2525, 'Z', 2527-2528, 'Z', 2530-2531, 'Z', 2533-2534, 'Z', 2536-2537, 'Z', 2539-2540, 'Z', 2542-2543, 'Z', 2545-2546, 'Z', 2548-2549, 'Z', 2551-2552, 'Z', 2554-2555, 'Z', 2557-2558, 'Z', 2560-2561, 'Z', 2563-2564, 'Z', 2566-2567, 'Z', 2569-2570, 'Z', 2572-2573, 'Z', 2575-2576, 'Z', 2578-2579, 'Z', 2581-2582, 'Z', 2584-2585, 'Z', 2587-2588, 'Z', 2590-2591, 'Z', 2593-2594, 'Z', 2596-2597, 'Z', 2599-2600, 'Z', 2602-2603, 'Z', 2605-2606, 'Z', 2608-2609, 'Z', 2611-2612, 'Z', 2614-2615, 'Z', 2617-2618, 'Z', 2620-2621, 'Z', 2623-2624, 'Z', 2626-2627, 'Z', 2629-2630, 'Z', 2632-2633, 'Z', 2635-2636, 'Z', 2638-2639, 'Z', 2641-2642, 'Z', 2644-2645, 'Z', 2647-2648, 'Z', 2650-2651, 'Z', 2653-2654, 'Z', 2656-2657, 'Z', 2659-2660, 'Z', 2662-2663, 'Z', 2665-2666, 'Z', 2668-2669, 'Z', 2671-2672, 'Z', 2674-2675, 'Z', 2677-2678, 'Z', 2680-2681, 'Z', 2683-2684, 'Z', 2686-2687, 'Z', 2689-2690, 'Z', 2692-2693, 'Z', 2695-2696, 'Z', 2698-2699, 'Z', 2701-2702, 'Z', 2704-2705, 'Z', 2707-2708, 'Z', 2710-2711, 'Z', 2713-2714, 'Z', 2716-2717, 'Z', 2719-2720, 'Z', 2722-2723, 'Z', 2725-2726, 'Z', 2728-2729, 'Z', 2731-2732, 'Z', 2734-2735, 'Z', 2737-2738, 'Z', 2740-2741, 'Z', 2743-2744, 'Z', 2746-2747, 'Z', 2749-2750, 'Z', 2752-2753, 'Z', 2755-2756, 'Z', 2758-2759, 'Z', 2761-2762, 'Z', 2764-2765, 'Z', 2767-2768, 'Z', 2770-2771, 'Z', 2773-2774, 'Z', 2776-2777, 'Z', 2779-2780, 'Z', 2782-27

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA
A;Residues: 1187-1194, C', 1196-1220 <CH>

A;Cross-references: GB:M23213; NID:G340842; PIDN:AAB59363.1; PID:G499622

A;Note: mutant sequence from a patient with mild osteogenesis imperfecta

R;MacKellae, J.K.; Raasina, M.; Virta, A.; Vuorio, E.

Nucleic Acids Res. 16, 349, 1988

A;Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

Query Match 97.0%; Score 1149; DB 1; Length 1464;
Best Local Similarity 93.6%; Pred. No. 1.8e-56;
Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 1 GPP-----GEPPTGLPPTGCGSGRGGPGGAGVAGPKGPRGSGPGPA 48

DB 458 GPPGPAGBEGKRGARKEPPTGLPPTGCGSGRGGPGGAGVAGPKGPRGSGPGPA 517

QY 49 GPKSGPGAGRPGEGAGLPGKGLTSGPSGPGDGTGTPGAGDGRPGPPGPGARGQA 108

DB 518 GPKSGPGAGRPGEGAGLPGKGLTSGPSGPGDGTGTPGAGDGRPGPPGPGARGQA 577

QY 109 GVMGPPGPKGAAGEGPKAGRGVPPGAVGPGAGKDGAGAGQPPGAPAGERGEQGPA 168

DB 578 GVMGPPGPKGAAGEGPKAGRGVPPGAVGPGAGKDGAGAGQPPGAPAGERGEQGPA 637

QY 169 GSPGPGQLPGAPGPPGEGAGKPGEGQVPGDILGAPGSGPAG 208

DB 638 GSPGPGQLPGAPGPPGEGAGKPGEGQVPGDILGAPGSGARG 677

RESULT 2

CORTIS

collagen alpha 1(I) chain - rat (tentative sequence) (fragments)

C;Species: Rattus norvegicus (Norway rat)

C;Date: 13-Jul-1981 #sequence revision 13-Jul-1981 #text change 31-Mar-2000

C;Accession: A90559; A90552; A92029; A90353; A90566; A90357; A90362; A90379; A91209; A91

R;Bornstein, P. 63-71, 1969

A;Title: Comparative sequence studies of rat skin and tendon collagen. II. The absence of

A;Reference number: A90559; MUID:69155173; PMID:5777344

A;Contents: CNBR0 and CNBR1

A;Accession: A90559

A;Molecule type: protein

A;Residues: 1-19 <B01>

A;Experimental source: tendon

A;Note: sequences from skin and tendon appear to be identical

A;Note: the amino-terminal tetrapeptide may be removed by limited proteolysis during ext

R;Kang, A.H.; Bornstein, P.; Piez, K.A.

Biochemistry 6, 788-795, 1967

A;Title: The amino acid sequence of peptides from the cross-linking region of rat skin c

A;Reference number: A90552; MUID:67162268; PMID:5337886

A;Contents: CNBR1

A;Accession: A90552

A;Molecule type: protein

A;Residues: 5-19 <KAN>

A;Experimental source: skin

R;Bornstein, P.

J. Biol. Chem. 242, 2572-2574, 1967

A;Title: The incomplete hydroxylation of individual prolyl residues in collagen.

A;Reference number: A92029; MUID:67165368; PMID:4290711

A;Contents: CNBR2

A;Accession: A92029

A;Molecule type: protein

A;Residues: 20-55 <B02>

A;Experimental source: skin and tendon

R;Butler, W.T.; Ponds, S.D.

Biochemistry 10, 2076-2081, 1971

A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino a

A;Reference number: A90353; MUID:71263178; PMID:4327399

A;Contents: CNBR4

A;Accession: A90353

A;Molecule type: protein

A;Residues: 56-102 <B01>

A;Experimental source: skin

R;Butler, W.T.

Biochemistry 9, 44-50, 1970

A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. The cove

A;Reference number: A90566; MUID:70085124; PMID:5411206

A;Contents: CNBR5

A;Accession: A90566

A;Molecule type: protein

A;Residues: 103-139 <B02>

A;Experimental source: skin

R;Balian, G.; Click, E.M.; Bornstein, P.

Biochemistry 10, 4470-4478, 1971

A;Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxylan

A;Reference number: A90357; MUID:72136131; PMID:4335087

A;Contents: CNBR8

A;Accession: A90357

A;Molecule type: protein

A;Residues: 140-238 <B01>

A;Experimental source: skin

R;Balian, G.; Click, E.M.; Hermodson, M.A.; Bornstein, P.

Biochemistry 11, 3798-3806, 1972

A;Title: Structure of rat skin collagen alpha1-CB8. Amino acid sequence of the hydroxylan

A;Reference number: A90362; MUID:73006942; PMID:4342027

A;Contents: CNBR8

A;Accession: A90362

A;Molecule type: protein

A;Residues: 239-418 <B02>

A;Experimental source: skin

R;Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.

Biochemistry 13, 2946-2953, 1974

A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino ac

A;Reference number: A90379; MUID:74271984; PMID:4366532

A;Contents: CNBR3

A;Accession: A90379

A;Molecule type: protein

A;Residues: 419-567 <B03>

A;Experimental source: skin

R;Stoltz, M.; Timpl, R.; Furthmayr, H.; Kuehn, K.

Eur. J. Biochem. 37, 287-294, 1973

A;Title: Structural and immunogenic properties of a major antigenic determinant in neutra

A;Reference number: A91209; MUID:74011954; PMID:4126850

A;Contents: CNBR6

A;Accession: A91209

A;Molecule type: protein

A;Residues: 568-651 <ST1>

A;Experimental source: skin

A;Note: this region probably corresponds to positions 949-1032 of the alpha 1(I) chain

A;Note: the major antigenic determinant (of neutral salt-extracted rat skin collagen) in

R;Stoltz, M.; Timpl, R.; Kuehn, K.

FEBS Lett. 26, 61-65, 1972

A;Title: Non-helical regions in rat collagen alpha1-chain.

A;Reference number: A91385; MUID:73049495; PMID:4636751

A;Contents: CNBR6

A;Accession: A91385

A;Molecule type: protein

A;Residues: 651-671 <ST2>

A;Experimental source: skin

A;Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the sequenc

A;Note: this region (residues 651-671 above) probably corresponds to positions 1032-1052

C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C

ed and subsequently O-glycosylated.

C;Comment: The order of the nine CNBR peptides in the alpha 1(I) chain of rat skin colla

C;Comment: The complete chain contains 1052 residues.

C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C;Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxyl;

F;1/Modified site: blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #status

F;9/Modified site: allylsine (Lys) #status experimental

F;103/424,547/Binding site: carboxylate (Lys) (covalent) #status experimental

F;103/Modified site: 5-hydroxylysine (Lys) #status experimental

F;424,547/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Query Match 94.4%; Score 1118; DB 1; Length 671;

Best Local Similarity 90.0%; Pred. No. 4.9e-55;

```
Matches 198; Conservative 5; Mismatches 5; Indels 12; Gaps 1;
QY 1 GPP-----GEPGPTGLPDPGCGSGRPGADGVAGKPGAGERSGCPA 48
Db 256 GPPGAGBEGKRGARGEPGLPDPGCGSGRPGADGVAGKPGAGERSGCPA 355
QY 49 GPKGSPGAGRPGEAGLPGAKGLTSGPSGPGDGTGPPGAGQDGRPPGPPGARGQA 108
Db 356 GPKGSPGAGRPGEAGLPGAKGLTSGPSGPGDGTGPPGAGZBGRPPGPPGARGQA 415
QY 109 GVMGFPFGKGAAGPFGKAGRGVPPGAVGAPGKDGAGAGQAGPAGPAGRGEGQPA 168
Db 416 GVMGFPFGKGTAGBFGKAGRGVPPGAVGAPGKDGAGAGQAGPAGPAGRGEGQPA 475
QY 169 GSPGFQGLPGPAGPPGEGKFGEGQVPGDLGAPGSPGAG 208
Db 476 GSPGFQGLPGPAGPPGEGKFGEGQVPGDLGAPGSPGARG 515

RESULT 3
S21626
collagen alpha 1(I) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text_change 09-Jul-2004
C:Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626
R:Li, S.W.; Kullian, J.; Prockop, D.J.
Matrix Biol. 14, 593-595, 1994
A>Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I
A:Reference number: S57243
A:Accession: S57243
A:Molecule type: mRNA
A:Residues: 1-1453 <LIS>
A:Cross-references: UNIPROT:P11087; EMBL:U08020; NID:9470673; PIDN:AAA88912.1; PID:G4706
P:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A>Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A:Reference number: S16176; MUID:91274355; PMID:2054384
A:Accession: S16374
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1453 <MET>
A:Cross-references: EMBL:X57981; NID:950484; PIDN:CAA41046.1; PID:g50485
R:French, B.T.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A>Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
A:Reference number: A23982; MUID:86137403; PMID:3841523
A:Accession: A23982
A:Molecule type: mRNA
A:Residues: 518-1128 <PRE>
A:Cross-references: GB:M14423; NID:g192261; PIDN:AAA37333.1; PID:g192262
R:Monson, J.M.; Friedman, J.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A>Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for
A:Reference number: I49559; MUID:83141374; PMID:6298597
A:Accession: I49559
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 735-1130 <RES>
A:Cross-references: GB:M17491; NID:g192263; PIDN:AAA37334.1; PID:g192264
P:Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
A>Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads
A:Reference number: I49557; MUID:84170331; PMID:6324198
A:Accession: I49557
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-25 <RE2>
A:Cross-references: GB:K01688; NID:g192246; PIDN:AAA37330.1; PID:g553881
R:Fenton, S.P.; Lamande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A>Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A:Reference number: S39789; MUID:94092741; PMID:8268229
A:Accession: S39789
```

```
A:Molecule type: DNA
A:Residues: 1-80, 'E', 82-105, 'D', 107-185, 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-1
R:Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.
Mol. Cell. Biol. 14, 5950-5960, 1994
A>Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indirect
A:Reference number: I48300; MUID:94344105; PMID:8065328
A:Accession: I48300
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>
A:Cross-references: EMBL:X54876; NID:g50486; PIDN:CAA38657.1; PID:g50487
C:Genetics:
A:Gene: COL1A1
A:Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/3
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple helix
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-151/Domain: amino-terminal propeptide #status predicted <PRO>
F:30-89/Domain: von Willebrand factor type C repeat homology <WVC>
F:152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>
F:1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 94.3%; Score 1116; DB 2; Length 1453;
Best Local Similarity 90.5%; Pred. No. 1.1e-54;
Matches 199; Conservative 5; Mismatches 12; Gaps 1;
QY 1 GPP-----GEPGPTGLPDPGCGSGRPGADGVAGKPGAGERSGCPA 48
Db 447 GPPGAGBEGKRGARGEPGLPDPGCGSGRPGADGVAGKPGAGERSGCPA 506
QY 49 GPKGSPGAGRPGEAGLPGAKGLTSGPSGPGDGTGPPGAGQDGRPPGPPGARGQA 108
Db 507 GPKGSPGAGRPGEAGLPGAKGLTSGPSGPGDGTGPPGAGQDGRPPGPPGARGQA 566
QY 109 GVMGFPFGKGAAGPFGKAGRGVPPGAVGAPGKDGAGAGQAGPAGPAGRGEGQPA 168
Db 567 GVMGFPFGKGTAGBFGKAGRGVPPGAVGAPGKDGAGAGQAGPAGPAGRGEGQPA 626
QY 169 GSPGFQGLPGPAGPPGEGKFGEGQVPGDLGAPGSPGAG 208
Db 627 GSPGFQGLPGPAGPPGEGKFGEGQVPGDLGAPGSPGARG 666

RESULT 4
CCHHS
collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
C:Species: Gallus gallus (chicken)
C>Date: 12-Aug-1981 #sequence revision 06-Jul-1982 #text_change 31-Mar-2000
C:Accession: A90458; A90181; A02857
R:Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.
Biochemistry 21, 2048-2055, 1982
A>Title: Amino acid sequence of chick skin collagen alpha1(I)-CB8 and the complete primary
A:Reference number: A90458; MUID:82231995; PMID:7093229
A:Accession: A90458
A:Molecule type: protein
A:Residues: 1-1036 <HIG>
A:Experimental source: skin
A>Note: This is the latest in a series of papers from these workers elucidating the sequ
R:Eyre, D.R.; Glimcher, M.J.
Biochem. Biophys. Res. Commun. 48, 720-726, 1972
A>Title: Evidence for a previously undetected sequence at the carboxyterminus of the alpt
A:Reference number: A90181; MUID:72243016; PMID:5047697
A:Accession: A90181
A:Molecule type: protein
A:Residues: 1037-1042 <EYR>
A:Experimental source: skin
A>Note: residues 1037-1042 above correspond to the carboxyl end of the protein
C:Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some
C:Comment: Most of the prolines at the third position of the tripeptide repeating unit (C
C:Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in pos
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimet;
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
```


A;Residues: 752-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'Q', 1038-1052, 'E', 1054-1068, 'T', 'A';
A;Cross-references: GB:L00977; NID:g180812; PIDN:AA23914.1; PID:g258774
A;Note: sequence extracted from NCBI backbone (NCBIP:117273); parts of this sequence were
A;Note: this translation is not annotated and this publication is not cited in GenBank
R;Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
A;Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual
A;Reference number: S16502; MUID:90251662; PMID:2339128
A;Accession: S16502
A;Molecule type: DNA
A;Residues: 1164-1184, 'GPSKDGANGIPQPI', 1185-1199, '<TII2>
A;Cross-references: EMBL:M37126; NID:g180808; PIDN:AAA52037.1; PID:g180809
A;Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
R;Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
A;Title: Identification and characterization of the human type II collagen gene (COL2A1)
A;Reference number: A02858; MUID:85190534; PMID:3857598
A;Accession: A02858
A;Molecule type: DNA
A;Residues: 1032-1056, 'N', 1058-1068, 'T', 1070-1487, '<CHE>
A;Cross-references: GB:J00116; NID:g180395; PIDN:AAA51997.1; PID:g180396
R;Elima, K.; Vuorio, T.; Vuorio, E.
Nucleic Acids Res. 15, 9499-9504, 1987
A;Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) c
A;Reference number: A27280; MUID:88067711; PMID:2825137
A;Accession: A27280
A;Molecule type: DNA; mRNA
A;Residues: 1175-1487, '<ELI>
A;Cross-references: EMBL:X06268; NID:g30096; PIDN:CAA29604.1; PID:g30097
A;Experimental source: fetal epiphyseal cartilage
R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
Biochem. J. 217, 923-925, 1986
A;Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
A;Reference number: A57033; MUID:87099927; PMID:3800925
A;Accession: A57033
A;Molecule type: protein
A;Residues: 'XE', 1244-1246, 'N', 1248, 'X', 1250-1265, 1295-1305, 1395-1408, '<VAN>
A;Note: chondrocalcin identified as released collagen I(II) chain carboxyl-terminal prope
R;Strom, C.M.; Upholt, W.B.
Nucleic Acids Res. 12, 1025-1038, 1984
A;Title: Isolation and characterization of genomic clones corresponding to the human type
A;Reference number: A21733; MUID:84118798; PMID:6320112
A;Accession: A21733
A;Molecule type: DNA
A;Residues: 1245-1295, '<STRI>
A;Cross-references: EMBL:X00339; EMBL:X00298; NID:g394699; PIDN:CAA25092.1; PID:g4378975
A;Accession: B21733
A;Molecule type: DNA
A;Residues: 894-909, 'PE', '<STR2>
A;Cross-references: GB:K01785; NID:g30035; PIDN:CAA25082.1; PID:g1335032
R;Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A;Title: Isolation and partial characterization of genomic clones coding for a human pro
gene.
A;Reference number: A24561; MUID:86104139; PMID:3002437
A;Accession: A24561
A;Molecule type: DNA
A;Residues: 1296-1358, '<NUN2>
A;Cross-references: GB:M12048; NID:g180017
A;Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
A;Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with the
R;Sangioi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez,
Nucleic Acids Res. 13, 2207-2225, 1985
A;Title: Isolation and partial characterization of the entire human pro alpha 1(II) coll
A;Reference number: I37249; MUID:85215609; PMID:2987845
A;Accession: S59491
A;Molecule type: DNA
A;Residues: 7-28; 'R', 99-114; 541-578; 786-802; 1055-1056, 'N', 1058-1068, 'T', 1070-1109; 1200-1
A;Accession: I84453
A;Status: translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 7-28, '<SAN2>

A;Cross-references: GB:M23759; NID:g180845; EMBL:X03320; GB:M24938; NID:g30104
A;Note: the GenBank PID is based on an incorrect reading frame
A;Accession: I37250
A;Status: translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 541-560, '<SAN3>
A;Cross-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929621
A;Accession: I37251

Query Match 72.8%; Score 862; DB 1; Length 1487;
Best Local Similarity 74.0%; Pred. No. 1.1e-40;
Matches 154; Conservative 11; Mismatches 43; Indels 0; Gaps 0;

QY 1 GPPGPGPTGLPGPPGPGGSGRCPGADGVAGPKGAGRGSPGPPAGPKSGPGEAGRP 60
Db 482 GARGPFGVGPIPPGPGRGAGPGRGFPQDGLAGPKGAPGRGSGLAGPKGANGDPGR 551

QY 61 GEAGLPGAKGLTSGSPGPGDKTGPAGQDGRPGGPPGARGQAGVMGFFPKGAA 120
Db 552 GEPGLPGARGLTGRPGDAGPOGKVGPSGAPGEDGRPGPGQARGQGVMGFFPKGAN 611

QY 121 GEPKAGRGVPGPGVAGVAGPKGDEAGAGQAPGPPGAGRGGQGPAGSPGQGLPGPA 180
Db 612 GEPKAGRGVPGPGVAGVAGPKGDEAGAGQAPGPPGAGRGGQGPAGSPGQGLPGPP 671

QY 181 GPPGAGKPGGQGVPGDGLGAPGSPGAG 208
Db 672 GPPGEGKPGQGVPGGAGAPGLVGRG 699

RESULT 6
T45467
N;Alternate names: type II chain precursor [imported] - horse
C;Species: Equus caballus (domestic horse)
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
C;Accession: T45467
R;Richardson, D.W.; Dodge, G.R.
A;Description: Cloning of equine type II collagen and modulation of its expression in eq
A;Reference number: 222977
A;Accession: T45467
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: mRNA
A;Residues: 1-1418, '<RIC>
A;Cross-references: UNIPROT:Q28396; EMBL:U62528; PIDN:AA805773.1
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 72.6%; Score 860; DB 2; Length 1418;
Best Local Similarity 74.0%; Pred. No. 1.3e-40;
Matches 154; Conservative 10; Mismatches 44; Indels 0; Gaps 0;

QY 1 GPPGPGPTGLPGPPGPGGSGRCPGADGVAGPKGAGRGSPGPPAGPKSGPGEAGRP 60
Db 423 GARGPFGVGPIPPGPGRGAGPGRGFPQDGLAGPKGAPGRGSGLAGPKGANGDPGR 482

QY 61 GEAGLPGAKGLTSGSPGPGDKTGPAGQDGRPGGPPGARGQAGVMGFFPKGAA 120
Db 483 GEPGLPGARGLTGRPGDAGPOGKVGPSGAPGEDGRPGPGQARGQGVMGFFPKGAN 542

QY 121 GEPKAGRGVPGPGVAGVAGPKGDEAGAGQAPGPPGAGRGGQGPAGSPGQGLPGPA 180
Db 543 GEPKAGRGVPGPGVAGVAGPKGDEAGAGQAPGPPGAGRGGQGPAGSPGQGLPGPP 602

QY 181 GPPGAGKPGGQGVPGDGLGAPGSPGAG 208
Db 603 GPPGEGKPGQGVPGGAGAPGLVGRG 630

RESULT 7
B40333
N;Alternate names: type II chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)

R.Fietzek, P.P.; Rexrodt, F.W.; Wendt, P.; Stark, M.; Kuehn, K.
Eur. J. Biochem. 30, 163-168, 1972
A:Title: The covalent structure of collagen. Amino acid sequence of peptide alpha1-CB6-807
A:Reference number: A91200; MUID:73042275; PMID:4343807
A:Accession: A91200
A:Molecule type: protein
A:Residues: 676-758 <FI4>
A:Experimental source: skin
A:Note: Pro-726 is the only 3-hydroxyproline and the only hydroxylated proline in position 726.
A:Author: Rauterberg, J.; Fietzek, P.; Rexrodt, F.; Becker, U.; Stark, M.; Kuehn, K.
FEBS Lett. 21, 75-79, 1972
A:Title: The amino acid sequence of the carboxyterminal nonhelical cross link region of type I collagen.
A:Reference number: A43048
A:Accession: A43048
A:Molecule type: protein
A:Residues: 759-779 <RA2>
A:Experimental source: skin
C:Comment: Lysines at positions 115, 124, 274, 346, 424, 496, 658, and 670 may be hydroxylated.
C:Comment: Prolines in the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated.
C:Comment: The order of the eight CNBr peptides in the alpha 1(I) chain of bovine skin collagen is: 1, 149, 268, and 217 residues.
C:Comment: The complete chain contains 1052 residues.
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimer; type I/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 59.6%; Score 824; DB 1; Length 779;
Best Local Similarity 72.0%; Pred. No. 7.9e-39;
Matches 154; Conservative 4; Mismatches 50; Indels 6; Gaps 1;

QY 1 GPPGEPGPTGLPGPPGGERGGPGS-----RGFGAGDGVAGPKGAGRGSGPPGAPGKSP 54
DB 29 GPGFAGPGQFGQPPGEPGFGSGGPMGPRGPPGPKNGDGDGAGKPRGFRGGRGPPGQ 88
QY 55 GEMRGEGEAGLPGAKGLTSGPSGPDGKTGPPGAGQDGRPGPPGPPGARGQAGVMGFP 114
DB 89 GARGLGFTAGLPGMKGHRGFGSLDGAKGDAGPAGPKGPGSGPENGAPGQMPGLPGFP 148
QY 115 GPKGAAGEPKAGERGVPCCPGAVGAGKDGAGAGQPPGAPGARGGOGGAGSGPGFQ 174
DB 149 GPKGAAGEFKAGERGVPCCPGAVGAGKDGAGAGQPPGAPGARGGOGGAGSGPGFQ 208
QY 175 GLPGPAGPPGEGAKGPGEGQGVPGDGLGAPGSPGAG 208
DB 209 GLPGPAGPPGEGAKGPGEGQGVPGDGLGAPGSPGAG 242

RESULT 12
CSG06C
collagen alpha 1(II) chain precursor - bovine (tentative sequence) (fragments)
C:Species: Bos primigenius taurus (cattle)
C:Date: 24-Apr-1984 #sequence revision 17-May-1996 #text change 09-Jul-2004
C:Accession: A90369; A90396; A92210; S03940; A90189; A05039; A02859
R:Miller, E.J.; Lunde, L.G.
Biochemistry 12, 3153-3159, 1973
A:Title: Isolation and characterization of the cyanogen bromide peptides from the alpha 1(II) chain of collagen.
A:Reference number: A90369; MUID:73258693; PMID:4732855
A:Contents: composition of CNBr1 and CNBr4
A:Accession: A90369
A:Molecule type: protein
A:Residues: 1-15 <MIL>
A:Cross-references: UNIPROT:P02459
A:Experimental source: cartilage
A:Note: residues positioned by comparison with human alpha 1(II) chain
R:Butler, W.T.; Miller, E.J.; Finch Jr., J.E.
Biochemistry 15, 3000-3006, 1976
A:Title: The covalent structure of cartilage collagen. Amino acid sequence of the NH-2-terminal peptide.
A:Reference number: A90396; MUID:76258504; PMID:782511
A:Contents: fragments CNBr2 (16-18), CNBr3 (19-21), CNBr6 (22-54), CNBr12 (55-138), and type I/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

A>Note: order of CNBr peptides determined
R;Butler, W.T.; Finch Jr., J.E.; Miller, E.J.
J. Biol. Chem. 252, 639-643, 1977
A>Title: The covalent structure of cartilage collagen. Evidence for sequence heterogeneity
A;Reference number: A92210; MUID:77093864; PMID:833147
A;Accession: A92210
A:Molecule type: protein
A;Residues: 139-178, 'Z', 180-184, 'PA', 187-190, 'AS', 193-194, 'T', 196-198 <BU2>
A;Experimental source: cartilage
A>Note: a minor, probably nonallelic, alpha 1(II) component has 143-Ala, 164-Leu, and proline at position 181.
R;Seyer, J.M.; Hasty, K.A.; Kang, A.H.
Eur. J. Biochem. 181, 159-173, 1989
A>Title: Covalent structure of collagen. Amino acid sequence of an arthritogenic cyanogen bromide peptide
A;Reference number: S03940; MUID:82231683; PMID:2714276
A;Accession: S03940
A:Molecule type: protein
A;Residues: 139-417 <SEY>
R;Butler, W.T.; Miller, E.J.; Finch Jr., J.E.; Inagami, T.
Biochem. Biophys. Res. Commun. 57, 190-195, 1974
A>Title: Homologous regions of collagen alpha 1(I) and alpha 1(II) chains: apparent cluster
A;Reference number: A90189; MUID:74163168; PMID:4857180
A;Accession: A90189
A:Molecule type: protein
A;Residues: 418-492 <BU3>
A;Experimental source: cartilage
A>Note: the first 75 residues of CNBr8, which follows CNBr11
R;Sangiorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Ramirez, F.
Nucleic Acids Res. 13, 2815-2826, 1985
A;Reference number: A05039; MUID:85215651; PMID:2582365
A;Accession: A05039
A:Molecule type: mRNA
A;Residues: 493-673 <SAN>
A;Cross-references: GB:X02420; NID:g265; PIDN:CAA26269.1; PID:g266
A;Experimental source: cartilage
C;Comment: Type II collagen molecules are trimers of identical alpha 1(II) chains, generated by the cleavage of the triple helix. The CNBr peptides were determined as 1-4-2-3-6-12-11-8-10-5-9-7-14-15.
C;Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology; C;Keywords: cartilage; coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline; collagen; carboxyl-terminal propeptide (fragment) #status predicted <CTP>
F;493-673/Domain: fibrillar collagen carboxyl-terminal homology (fragment) <FCC>
F;9,102,114,123,189,423,435/Modified site: 5-hydroxylysine (Lys) #status experimental
F;9,102,114,123,189,423,435/Binding site: carboxylate (Asn) (covalent) #status experimental
F;574/Binding site: carboxylate (Asn) (covalent) #status predicted
Query Match 68.8%; Score 815; DB 1; Length 673;
Best Local Similarity 78.9%; Pred. No. 2.2e-38;
Matches 146; Conservative 4; Mismatches 35; Indels 0; Gaps 0;
Qy 1 GPPGPGTGLPGRGPGSGRFGPGADGVAGPKGPGAGRGSPGPGKSPGAGRP 60
Db 307 GARGEPGAGPAGPPGERGAPGRGFPQDGIAGPKGPPGRGSPGAVGPKGSPGAGRP 366
Qy 61 GEGLPGAKGLTSGSPGPGKTPGPGAGQDGRPGPPGARGQAGVGFPGKGA 120
Db 367 GEGLPGAKGLTSGSPGPGKTPGPGAGQDGRPGPPGARGQAGVGFPGKGA 426
Qy 121 GEPKAGRGVPPGPPGAVGPGAGKDGAGAGQAGQPGAGRGQGGAGSPGFGQLPGPA 180
Db 427 GEPKAGRGVPPGPPGAVGPGAGKDGAGAGQAGQPGAGRGQGGAGSPGFGQLPGPA 486
Qy 181 GPPGE 185
Db 487 GPPGE 491
RESULT 13
150694
collagen alpha 1(III) chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
A;Accession: 150694
R;Nah, H.D.; Niu, Z.; Adams, S.L.

J. Biol. Chem. 269, 16443-16448, 1994
A>Title: An alternative transcript of the chick type III collagen gene that does not encode the triple helix
A;Reference number: A54041; MUID:94266842; PMID:8206952
A;Accession: 150694
A;Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A;Residues: 1-886 <NAH>
A;Cross-references: UNIPROT:P12105; EMBL:U07973; NID:g520454; PIDN:AAA83407.1; PID:g5374;
C;Genetics:
A;Gene: COL3A1
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; F;30-90/Domain: von Willebrand factor type C repeat homology <VMC>
Query Match 67.1%; Score 794.5; DB 2; Length 886;
Best Local Similarity 65.9%; Pred. No. 3.6e-37;
Matches 143; Conservative 13; Mismatches 52; Indels 9; Gaps 1;
Qy 1 GPP-----GPPGPTGLPGRGPGSGRFGPGADGVAGPKGPGAGRGSPGAGPK 51
Db 458 GPGEKRGKANGEPGQGVGTFGEGSGFGFGLPGNSGLPGEKGPAGRGSGFGPGPS 517
Qy 52 GSPGEAGRGPEAGLPGAKGLTSGSPGPGDKTGPAGQDGRPGPPGARGQAGVM 111
Db 518 GPAGDRQDGGPGLPGMRGLPGIPGSPGSGKPGPGNQGEPGSGPGAGPRGPGVM 577
Qy 112 GFPGPKAAGEPKAGRGVPGPGAVGPGAGKDGAGAGQAGQPGAGRGEGQPGASP 171
Db 578 GFPGPKNEGAPKNGRGPGGPGTGTGAGKNDGVDGLPGPPGAGPRGEGFGSGSP 637
Qy 172 GFQGLPGAGPAGGAGKPGGQGVPGDLGAPGSPGAG 208
Db 638 GLQGLPGGPGAGENGKPGGPKGDI GGFGFGPKG 674
RESULT 14
CGHU2V
collagen alpha 2(V) chain precursor - human
C;Species: Homo sapiens (man)
C;Date: 31-Jul-1989 #sequence_revision 28-Jul-1995 #text_change 09-Jul-2004
A;Accession: A31427; A54555; S43643; A25874; I59235; A25374; A30017
R;Woodbury, D.; Benson-Chanda, V.; Ramirez, F.
J. Biol. Chem. 264, 2735-2738, 1989
A>Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the structure of the alpha2(V) chain
A;Reference number: A31427; MUID:89123368; PMID:2914927
A;Accession: A31427
A:Molecule type: mRNA
A;Residues: 1-463 <WOO>
A;Cross-references: UNIPROT:P05997; GB:J04478; NID:g179697; PIDN:AAA51859.1; PID:g179698
A;Experimental source: placenta
R;Greenspan, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.
Gene Expr. 1, 29-39, 1991
A>Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence for a common regulatory region
A;Reference number: A54555; MUID:92314691; PMID:1820205
A;Accession: A54555
A:Molecule type: DNA
A;Residues: 1-32 <GRE>
A;Cross-references: GB:M58529; NID:g180834; PIDN:AAC41699.1; PID:g553235
R;Moradi-Ameli, M.; Rousseau, J.C.; Kleman, J.P.; Champilaud, M.F.; Boutillon, M.M.; Berr
Eur. J. Biochem. 221, 987-995, 1994
A>Title: Diversity in the processing events at the N-terminus of type-V collagen.
A;Reference number: S43642; MUID:94237164; PMID:8181482
A;Accession: S43642
A:Molecule type: protein
A;Residues: 288-291, 'P', 293-294, 'X', 296-297, 606, 'X', 608-617 <MOR>
R;Weil, D.; Bernard, M.; Gargano, S.; Ramirez, F.
Nucleic Acids Res. 15, 181-198, 1987
A>Title: The pro alpha 2(V) collagen gene is evolutionarily related to the major fibrillar collagen gene
A;Reference number: A25874; MUID:87146331; PMID:3029669
A;Accession: A25874
A:Molecule type: mRNA; DNA
A;Residues: 398-1496 <WEI>
A;Cross-references: GB:X04758; NID:g295588; PIDN:CAA28454.1; PID:g1340175
A;Experimental source: Rhabdomyosarcoma cell line

R;Myers, J.C.; Loidl, H.R.; Stollie, C.A.; Seyer, J.M.
 J. Biol. Chem. 260, 5533-5541, 1985
 A;Title: Partial covalent structure of the human alpha 2 type V collagen chain.
 A;Reference number: I55239; MUID:85182703; PMID:298598
 A;Accession: I55239
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1002-1226 <RE2>
 A;Cross-references: GB:M10956; NID:g180427; PIDN:AAA52007.1; PID:g180428
 A;Note: Part of this sequence were determined by protein sequencing
 R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
 A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
 A;Reference number: I59025; MUID:85216505; PMID:3859826
 A;Accession: I59025
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1003-1034 <RES>
 A;Cross-references: GB:M1135; NID:g179693; PIDN:AAA51857.1; PID:g179694
 A;Note: Part of this sequence were determined by protein sequencing
 R;Myers, J.C.; Loidl, H.R.; Seyer, J.M.; Dion, A.S.
 J. Biol. Chem. 260, 11216-11222, 1985
 A;Title: Complete primary structure of the human alpha-2 type V procollagen COOH-terminal
 A;Reference number: A25374; MUID:85289337; PMID:2411731
 A;Accession: A25374
 A;Molecule type: mRNA
 A;Residues: 1227-1417, 'T', 1419-1437, 'S', 1439-1496 <MYE>
 A;Cross-references: GB:M11718; NID:g180912; PIDN:AAA52058.1; PID:g180913
 A;Experimental source: normal fibroblasts
 R;Tsipouras, P.; Schwartz, R.C.; Liddell, A.C.; Salkeld, C.S.; Weil, D.; Ramirez, F.
 Genomics 3, 275-277, 1988
 A;Title: Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2, located on
 A;Reference number: A30017; MUID:89138450; PMID:3224983
 A;Accession: A30017
 A;Molecule type: DNA
 A;Residues: 1449-1463, 'E', 1465-1495, 'A' <TSI>
 A;Cross-references: GB:J03051; NID:g179695; PIDN:AAA51858.1; PID:g179696
 A;Note: The authors translated the codon GAA for residue 1460 as Gln, and GAG for residue
 C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 are 5-hydroxylated and subsequently O-glycosylated.
 C;Comment: The amino-terminal propeptide domain appears not to be completely cleaved.
 C;Genetics:
 A;Gene: GDB:COL5A2
 A;Cross-references: GDB:113064; OMIM:120190
 A;Map position: 2q31-2q31
 A;Introns: 33/1; 812/3; 830/3; 848/3; 884/3; 902/3; 922/3; 974/3; 1046/3; 1064/3; 1448/3
 A;Complex: type V collagen may be a homotrimer of alpha 1(V) chains (see PIR:CGHUIV), a
 alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the
 engh, is formed with desmosine cross-links made from lysine and allysine residues
 C;Function:
 A;Description: structural component of extracellular fibrous polymer associated with cell
 A;Note: may play a role in controlling the lateral growth of collagen I fibrils
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C;Keywords: coiled coil, extracellular matrix; glycoprotein; hydroxylysine; hydroxyproli
 F;1-26/Domain: signal sequence #status predicted <SIG>
 F;27-1250/Product: collagen alpha 2(V) chain #status predicted <MAT>
 F;27-193/Domain: amino-terminal propeptide (uncleaved) #status predicted <NPP>
 F;40-98/Domain: amino-terminal factor type C repeat homology <VWC>
 F;109-186/Region: von Willebrand factor type C repeat homology
 F;187-208/Region: helical
 F;209-1225/Region: nonhelical
 F;503-505/Region: cell attachment (R-G-D) motif
 F;941-943/Region: cell attachment (R-G-D) motif
 F;1064-1066/Region: cell attachment (R-G-D) motif
 F;1067-1069/Region: cell attachment (R-G-D) motif
 F;1097-1099/Region: cell attachment (R-G-D) motif
 F;1124-1126/Region: cell attachment (R-G-D) motif
 F;1133-1135/Region: cell attachment (R-G-D) motif
 F;1225-1250/Region: carboxyl-terminal nonhelical
 F;1251-1496/Domain: carboxyl-terminal propeptide #status predicted <CPP>
 F;1269-1496/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 F;27/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F;193-194/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted
 F;194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
 F;201/Modified site: allysine (Lys) #status predicted
 F;290,293,296,608,614,1004,1007,1013,1028,1034/Modified site: 4-hydroxyproline (Pro) #sta
 F;299,1139/Modified site: 5-hydroxylysine (Lys) #status predicted
 F;299,1139/Binding site: carboxylate (Lys) #status predicted
 F;1025/Modified site: 5-hydroxylysine (Lys) (covalent) #status predicted
 F;1250-1251/Cleavage site: Glu-Asp (procollagen C-endopeptidase) #status predicted
 F;1259,1397/Binding site: carboxylate (Asn) (covalent) #status predicted
 F;1293,1299,1325/Disulfide bonds: interchain #status predicted
 F;1333-1494,1402-1447/Disulfide bonds: #status predicted

Query Match 66.4%; Score 786; DB 1; Length 1496;
 Best Local Similarity 67.3%; Pred. No. 1.6e-36;
 Matches 140; Conservative 15; Mismatches 53; Indels 0; Gaps 0;

QY	1	GPPEPGPTGLPPGPPGSGRFGPCADGVAGKGPAGRGSPGAGPKGSGPEAGRP	60
DB	501	GPFGDPTGLPPGPPGSGRFGPCADGVAGKGPAGRGSPGAGPKGSGPEAGRP	560
QY	61	CEAGLPGAKGLTGTSPGSPDGKTPPGAGQDGRGPPGPPGARGQAGVGMFPKPGAA	120
DB	561	CEPGLPGARGLTGNPGVQGPBGKLGPLGAPGEDGCRPGPGSIGIKGQGTWGLPGKGSN	620
QY	121	CEPGKAGRGVPPGPGVAGVAGKAGAGAGGPGGPGAGRGSPGAGPKGSGPEAGRP	180
DB	621	GDPKFGEAGNPGVPGRGAPGKDGKVGYPGPPGPPGLRGERGQGGPPGPTGFGHPGPP	680
QY	191	GPPEAGKPGQGVPGDLGAPGSPGAP	208
DB	681	GPPEAGKPGQGVPGDLGAPGSPGAP	708

RESULT 15

CGHUIV

collagen alpha 1(III) chain precursor - human

N;Alternate names: procollagen alpha 1(III) chain

C;Species: Homo sapiens (man)

C;Date: 24-Apr-1994 #sequence revision 01-Sep-1995 #text change 09-Jul-2004

C;Accession: S05272; S04642; F8011; S01726; S04887; A90399; A94562; I51868; S59511; A904

R;Prockop, D.J.

submitted to the EMBL Data Library, February 1989

A;Reference number: S05272

A;Accession: S05272

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1240, 'V', 1242-1466 <PRC>

A;Cross-references: UNIPROT:P02461; EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058

R;Ala-Kokko, J.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.

Biochem. J. 260, 509-516, 1989

A;Title: Structure of cDNA clones coding for the entire prepro-alpha(I) chain of human

erences.

A;Reference number: S04642; MUID:89350838; PMID:2764886

A;Accession: S04642

A;Molecule type: mRNA

A;Residues: 1-1196 <ALA>

A;Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058

A;Note: the complete sequence is not shown

R;Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.

Gene 78, 255-265, 1989

A;Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene ((

A;Reference number: F8011; MUID:89378752; PMID:2777083

A;Accession: F8011

A;Molecule type: DNA

A;Residues: 1-176 <BEN>

A;Cross-references: GB:M26939; NID:g180813; PIDN:AAA52040.1; PID:g180814

R;Toman, P.D.; Ricca, G.A.; de Crombrughe, B.

Nucleic Acids Res. 16, 7201, 1988

A;Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pref

A;Reference number: S01726; MUID:88303360; PMID:3405773

A;Accession: S01726

A;Molecule type: mRNA

A;Residues: 1-170 <TOM>

A;Cross-references: EMBL:X07240; NID:G30060; PIDN:CAA30229.1; PID:G30061
A;Note: the authors translated the codon CAG for residue 154 as His
R;Janeczko, R.A.; Ramirez, F.
A;Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.
A;Reference number: S04887; MUID:89386015; PMID:2780304
A;Accession: S04887
A;Molecule type: mRNA
A;Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634,
A;Cross-references: EMBL:X15332; NID:G29545; PIDN:CAA33387.1; PID:G930045
A;Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
R;Seyer, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A;Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A;Reference number: A90399; MUID:77134724; PMID:557335
A;Accession: A90399
A;Molecule type: protein
A;Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEV1>
A;Experimental source: liver
A;Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose
R;Seyer, J.M.
submitted to the Atlas, December 1977
A;Reference number: A94562
A;Accession: A94562
A;Molecule type: protein
A;Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEV2>
A;Experimental source: liver
A;Note: author submitted corrections to A90399
R;Milewicz, D.M.; Witt, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A;Title: Parental somatic and germ-line mosaicism for a multigene deletion with unusual
tefring.
A;Reference number: I51868; MUID:93304430; PMID:8317500
A;Accession: I51868
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 186-194 <ML>
A;Cross-references: GB:S62925; NID:G386425; PIDN:AA013937.1; PID:G4261637
R;Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A;Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3A1
A;Reference number: S95511; MUID:96067614; PMID:7487954
A;Accession: S95511
A;Molecule type: mRNA
A;Residues: 302-423 <CHI>
A;Cross-references: GB:S79877; NID:G1195576; PIDN:AA035615.1; PID:G1195577
R;Seyer, J.M.; Kang, A.H.
Biochemistry 17, 3404-3411, 1978
A;Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr peptides
A;Reference number: A90414; MUID:79000343; PMID:687591
A;Accession: A90414
A;Molecule type: protein
A;Residues: 399-675, 'N', 677-727 <SEV3>
A;Experimental source: liver
A;Note: sequence corrected by A94562
R;Lee, B.; Vitale, E.; Superi-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991
A;Title: G to T transversion at position +5 of a splice donor site causes skipping of the
A;Reference number: I55349; MUID:91161621; PMID:1672129
A;Accession: I55349
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 537-605 <LEE>
A;Cross-references: GB:M59312; NID:G180815; PIDN:AAA52041.1; PID:G180816
R;Seyer, J.M.; Mainardi, C.; Kang, A.H.
Biochemistry 19, 1583-1589, 1980
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from type
A;Reference number: A90438; MUID:80198282; PMID:6246925
A;Accession: A90438
A;Molecule type: protein
A;Residues: 728-895, 'A', 897-964 <SEV4>
A;Experimental source: liver
R;Cole, W.G.; Chiodo, A.A.; Lamanche, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan
J. Biol. Chem. 265, 17070-17077, 1990

A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping and
A;Reference number: A98303; MUID:91009133; PMID:2145268
A;Accession: A98303
A;Molecule type: mRNA
A;Residues: 861-1015 <COL>
A;Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:G180878; PIDN:AA059383.1; PID:G1
A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syndrome
R;Mankoo, B.S.; Dalglish, R.
Nucleic Acids Res. 16, 2337, 1988
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A;Reference number: S02119; MUID:88189827; PMID:3357782
A;Accession: S02119
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>
A;Cross-references: EMBL:X06700; NID:G30053; PIDN:CAA29886.1; PID:G30054
R;Seyer, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1991
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from type
A;Reference number: A90446; MUID:81208139; PMID:7016180
A;Accession: A90446
A;Molecule type: protein
A;Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-1
A;Experimental source: liver
R;Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Meyer
Nucleic Acids Res. 12, 9383-9394, 1984
A;Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollagen
A;Reference number: A93551; MUID:85087944; PMID:6096827
A;Accession: A93551
A;Molecule type: mRNA
A;Residues: 1065-1155, 'P', 1157-1466 <LOI>
A;Cross-references: EMBL:X01655; EMBL:X01742; NID:G29584; PIDN:CAA25821.1
R;Miskulin, M.; Dalglish, R.; Kluge-Becker, B.; Rennard, S.I.; Tolstoshev, P.; Brant
Biochemistry 25, 1408-1413, 1986
A;Title: Human type III collagen gene expression is coordinately modulated with the type
A;Reference number: I52393; MUID:86187804; PMID:3754462
A;Accession: I52393
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1161-1200 <MIS>
A;Cross-references: GB:M13146; NID:G180415; PIDN:AAA52003.1; PID:G180416
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A;Reference number: I59025; MUID:85216505; PMID:3858826
A;Accession: I59025
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1165-1196 <EMA>
A;Cross-references: GB:M11134; NID:G180417; PIDN:AAA52004.1; PID:G180418
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A;Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. P.
A;Reference number: A92516; MUID:85157600; PMID:2579949
A;Accession: A92516
A;Molecule type: DNA
A;Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>
A;Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB:
A;Experimental source: liver
A;Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given for
ation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C
3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently O-
C;Genetics:
A;Gene: GDB:COL3A1
A;Map position: 2q31-2q31
A;Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Danlos
C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfide br
er of their length, is formed with desmosine cross-links made from lysine and allysine re
C;Function:
A;Description: structural component of extracellular fibrous polymer that maintains integ

Query Match	65.5%;	Score 776;	DB 1;	Length 1466;
Best Local Similarity	66.3%;	Pred. No. 5.5e-36;		
Matches 138;	Conservative 12;	Mismatches 58;	Indels 0;	Gaps 0;
Qy	1	GPFGFGPTGLPPGPPGGRGSGRFGGAGVAGPFGKPGAGERGSGPCGAPGKSPGCEAGRP	60	
Db	468	GSPFGGANGLPGAGRGACGFRGAPGNGIPGEXGPAGERGAPGAPRGAAGEPRD	527	
Qy	61	GEAGLPGAKGLTGGSPGSPDGKTPGPPAGQDGRFGPPGPPCARGAQGVGWFPGPKGAA	120	
Db	528	GVFGFGNGRMFGSPFGSGDKPFPSPGSGESGRFGPPGSPRGQFGVGWFFGPKGND	587	
Qy	121	GGPKKAGERGVGPPGAVGAGKDGAGAGCGPPGAPAGERGEGGPGSGFGQLPGPA	180	
Db	588	GAPGKNGRGCGFGPGGPPGKNGETGPPGPGTGGGDKGDTGPPGPGQLGQFGTG	647	
Qy	181	GGPGAGKPGGQGVPGDLAGPSPGAG	208	
Db	648	GPPGNGKGPFGPKGDAGAFAPGGKG	675	

Search completed: June 17, 2005, 15:15:12
Job time : 14.1543 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:17:54 ; Search time 31.0018 Seconds
(without alignments)
3452.202 Million cell updates/sec

Title: US-10-658-989A-1

Perfect score: 1184

Sequence: 1 GPPGEPGPTGLPGPPGRRG.....GEQVPGDLGAPGSPGAG 209

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1149	97.0	1069	2	Q61AN8
2	1149	97.0	1460	1	CA11_CANFA
3	1149	97.0	1461	2	O76045
4	1149	97.0	1464	1	CA11_HUMAN
5	1149	97.0	1464	2	Q8N473
6	1122	94.8	1453	2	Q63079
7	1118	94.4	671	1	CA11_RAT
8	1116	94.3	1225	2	Q6PCL3
9	1116	94.3	1453	1	CA11_MOUSE
10	1116	94.3	1453	2	Q810J9
11	1067	90.1	1453	1	CA11_CHICK
12	989	83.5	1450	2	Q9YIB4
13	981	82.9	1445	2	Q93251
14	974.5	82.3	1449	2	Q640B2
15	962.5	81.3	1449	2	Q802B5
16	941	79.5	1447	2	Q9IB91
17	913	77.1	747	1	CA12_BOVIN
18	862	72.8	1418	1	CA12_HUMAN
19	862	72.8	1447	2	Q6UIJ5
20	860	72.6	1418	2	Q28396
21	860	72.6	1487	2	O77753
22	859	72.6	1160	2	Q14046
23	859	72.6	1269	2	O7T227
24	859	72.6	1420	2	Q90W37
25	859	72.6	1487	2	Q14047
26	853	72.0	1447	2	Q6P4U1
27	852	72.0	1492	2	Q6P4Z2
28	850	71.8	1486	2	Q91717
29	850	71.8	1486	2	O7ZTI6
30	844	71.3	1491	2	Q91718
31	844	71.3	1491	2	Q7ZTM4

32	839	70.9	1419	2	Q63123
33	839	70.9	1442	2	Q62031
34	839	70.9	1442	2	Q62033
35	839	70.9	1459	1	CA12_MOUSE
36	839	70.9	1459	2	Q62032
37	835	70.5	1419	2	Q80VY3
38	835	70.5	1419	2	Q80X38
39	835	70.5	1487	2	Q641K3
40	833	70.4	1418	2	Q9W7K9
41	833	70.4	1449	2	Q910C0
42	824	69.6	779	1	CA11_BOVIN
43	800	67.6	1449	2	Q6NZI5
44	800	67.6	1449	2	Q6PEI9
45	798	67.4	1347	2	Q96QB3

ALIGNMENTS

RESULT 1

Q61AN8	PRELIMINARY;	PRT; 1069 AA.
AC	Q61AN8;	
DT	05-JUL-2004 (TrEMBLrel. 27, Created)	
DT	05-JUL-2004 (TrEMBLrel. 27, Last sequence update)	
DE	05-JUL-2004 (TrEMBLrel. 27, Last annotation update)	
DE	Collagen type I alpha 1 (fragment).	
GN	Name=COL1A1;	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=97141927; PubMed=8989177;	
RA	Simon M., Pedutour F., O'Brien K., Sirvent N., Kedra D., Guilbaud C.,	
RA	Turc-Carel C., Dumanski J.P.;	
RT	"Regulation of the platelet-derived growth factor B-chain gene via	
RT	fusion with collagen gene COL1A1 in dermatofibroblastoma."	
RT	and giant-cell fibroblastoma."	
RL	Nat. Genet. 15:95-98(1997).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RA	O'Brien K.P.;	
RL	EMBL; X88705; CAA67261.1; -.	
DR	GO; GO:0005737; C:cytoplasm; IEA.	
DR	GO; GO:0006817; P:phosphate transport; IEA.	
DR	InterPro; IPR008161; Clg_helix.	
DR	InterPro; IPR008160; Collagen.	
DR	InterPro; IPR009041; PMP_SGCI.	
DR	InterPro; IPR001007; VWF_C.	
DR	Pfam; PF01391; Collagen; 16.	
DR	Pfam; PF00093; VWC; 1.	
DR	ProDom; PD000007; Clg_helix; 3.	
DR	SMART; SM00214; VWC; 1.	
DR	PROSITE; PS01208; VWF_C_1; 1.	
DR	PROSITE; PS01084; VWF_C_2; 1.	
KW	Collagen.	
FT	NON_TER.	
SQ	SEQUENCE 1069 AA; 97445 MW; EE279B10572FB980 CRC64;	

Query Match 97.0%; Score 1149; DB 2; Length 1069;

Best Local Similarity 93.6%; Pred. No. 6.3e-44;

Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

Oy 1 GPP-----GEPGTGLPGPPGRRGPGSGRFPAGDVGAGPKGAGERGSPGA 48

Db 458 GPPGPAGEGKRGARGEPGTGLPGPPGRRGPGSGRFPAGDVGAGPKGAGERGSPGA 517

Oy 49 GPKSGPEAGRPAGLPGAKLTGSPGSPGDKGTGPPAGODGPPGPPGARGQA 108

Db 518 GPKSGPEAGRPAGLPGAKLTGSPGSPGDKGTGPPAGODGPPGPPGARGQA 577

FEBS Lett. 279:9-13 (1991).
 [5] SEQUENCE FROM N.A.
 RA MEDLINE=92157916; PubMed=1787829;
 RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
 RA Olsen A.S., Prockop D.J.;
 RA "Completion of the last half of the structure of the human gene for
 RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
 RL Matrix 11:375-379 (1991).
 [6]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=98107942; PubMed=9443882;
 RA Korkko J., Ala-Kokko L., De Paepae A., Nuytinck L., Earley J.,
 RA Prockop D.J.;
 RA "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
 RT scanning by conformation-sensitive gel electrophoresis identifies only
 RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I:
 RT identification of common sequences of null-allele mutations.";
 RL Am. J. Hum. Genet. 62:98-110 (1998).
 [7]
 RN SEQUENCE FROM N.A.
 RA Korkko J.M., Earley J.J., Nuytinck L., DePaepae A., Prockop D.J.,
 RA Ala-Kokko L.;
 RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AF017178; AAB94054.2;
 DR GO: GO:0005581; C:collagen; IEA.
 DR GO: GO:0005737; C:cytoplasm; IEA.
 DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR GO: GO:0006817; P:phosphate transport; IEA.
 DR InterPro: IPR008161; Clq helix.
 DR InterPro: IPR008160; Collagen.
 DR InterPro: IPR000885; Fib collagen C.
 DR InterPro: IPR009041; PMP SGC1.
 DR InterPro: IPR001007; VWF C.
 DR Pfam: PF01410; COLFI; 1.
 DR Pfam: PF01391; Collagen; 18.
 DR ProDom: PD000007; Clq_helix; 3.
 DR ProDom: PD002078; Fib_collagen_C; 1.
 DR SMART: SM00039; COLFI; 1.
 DR SMART: SM00214; VWC; 1.
 DR PROSITE: PS01208; VWF_C_1; 1.
 DR PROSITE: PS01184; VWF_C_2; 1.
 KW Collagen.
 SQ SEQUENCE 1461 AA; 138629 MW; 9ACF6D830EA78E21 CRC64;
 Query Match 97.0%; Score 1149; DB 2; Length 1461;
 Best Local Similarity 93.6%; Pred. No. 7.7e-44;
 Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GPGPTGLPDPGPGSGRGGPGGADGVAGPKGPGAGRGSGPGA 48
 DB |||||
 DB 455 GPPGAGGEGKRGARGEPTGLPDPGPGSGRGGPGGADGVAGPKGPGAGRGSGPGA 514
 QY 49 GPKSGPGAGPGGAGLPGKGLTSGSPGPDGKTGPPGAGDGRGPGPGPGCA 108
 DB |||||
 DB 515 GPKSGPGAGPGGAGLPGKGLTSGSPGPDGKTGPPGAGDGRGPGPGPGCA 574
 QY 109 GVMGPPGKGAAGEGKAGRGVPPGAGVAGPKDGAAGQGGPGGAGRGEGQGA 168
 DB |||||
 DB 575 GVMGPPGKGAAGEGKAGRGVPPGAGVAGPKDGAAGQGGPGGAGRGEGQGA 634
 QY 169 GSPGFGQLPGPAGPGGAGKPGGQGVGDGLGAPGSPGAG 208
 DB |||||
 DB 635 GSPGFGQLPGPAGPGGAGKPGGQGVGDGLGAPGSPGAG 674
 RESULT 4
 CALL HUMAN
 ID CALL HUMAN STANDARD; PRT: 1464 AA.
 AC P02452; P78441; Q13896; Q13902; Q13903; Q14037; Q14992; Q15176;
 AC Q15201; Q16050; Q7K230; Q7K234; Q8IV15; Q9UML6; Q9UML7;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)

25-JAN-2005 (Rel. 46, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN Name=COL1A1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT SER-1434.
 RA Dalglish R.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 [2]
 RN SEQUENCE OF 1-589 FROM N.A.
 RP MEDLINE=98329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
 RA D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,
 RA Pretorius P.J.;
 RA "Complete nucleotide sequence of the region encompassing the first
 RT twenty-five exons of the human pro alpha 1(I) collagen gene
 RT (COL1A1).";
 RL Gene 67:105-115 (1988).
 [3]
 RN SEQUENCE OF 1-472 FROM N.A.
 RP MEDLINE=89025644; PubMed=3178743;
 RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
 RA Jaenisch R., Prockop D.J.;
 RA "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
 RT of human type I procollagen.";
 RL Biochem. J. 253:919-922 (1988).
 [4]
 RN SEQUENCE OF 1-181 FROM N.A.
 RP MEDLINE=84270697; PubMed=6462220;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
 RA Myers J., Williams C., Ramirez F.;
 RA "Human pro alpha 1(I) collagen gene structure reveals evolutionary
 RT conservation of a pattern of introns and exons.";
 RL Nature 310:337-340 (1984).
 [5]
 RN SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.
 RP TISSUE=Skin;
 RC MEDLINE=71038625; PubMed=5529814;
 RA Click E.M., Bornstein P.;
 RA "Isolation and characterization of the cyanogen bromide peptides from
 RT the alpha 1 and alpha 2 chains of human skin collagen.";
 RL Biochemistry 9:4699-4706 (1970).
 [6]
 RN SEQUENCE OF 425-1464 FROM N.A.
 RP MEDLINE=84080385; PubMed=6689127;
 RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
 RA Prockop D.J.;
 RA "Nucleotide sequences of complementary deoxyribonucleic acids for the
 RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
 RT of structures that are conserved during evolution.";
 RL Biochemistry 22:5213-5223 (1983).
 [7]
 RN SEQUENCE OF 472-607 FROM N.A.
 RP PubMed=2981843;
 RA Chu M.-L., Gargiulo V., Williams C.J., Ramirez F.;
 RA "Multixon deletion in an osteogenesis imperfecta variant with
 RT increased type III collagen mRNA.";
 RL J. Biol. Chem. 260:691-694 (1985).
 [8]
 RN SEQUENCE OF 488-625 FROM N.A.
 RP PubMed=3857621;
 RA Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gelinas R.E.;
 RA "Intron-mediated recombination may cause a deletion in an alpha 1 type
 RT I collagen chain in a lethal form of osteogenesis imperfecta.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874 (1985).
 [9]
 RN SEQUENCE OF 1179-1464 FROM N.A., VARIANTS OI-II HIS-1277; ARG-1388 AND
 RP 1337-GLU-TYR-1338 DEL, AND VARIANT SER-1434.
 RX MEDLINE=93352646; PubMed=8349697;
 RA Chessier S.D., Wallis G.A., Byers P.H.;
 RA "Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I)

RT chain of type I collagen result in defective chain association and
 RT produce lethal osteogenesis imperfecta.";
 RL J. Biol. Chem. 268:18218-18225(1993).
 RN [10]
 RP SEQUENCE OF 1229-1454 FROM N.A.
 RC TISSUE=Bone;
 RX MEDLINE=88124208; PubMed=3340531;
 RA Maekelae J.K., Raassina M., Virta A., Vuorio E.;
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
 RT domain.";
 RL Nucleic Acids Res. 16:349-349(1988).
 RN [11]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=85130970; PubMed=2857713;
 RA Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;
 RT "Fine structural analysis of the human pro-alpha 1(I) collagen gene.
 RT Promoter structure, AluI repeats, and polymorphic transcripts.";
 RL J. Biol. Chem. 260:2315-2320(1985).
 RN [12]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=88097389; PubMed=3480516;
 RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
 RT "Regulatory elements in the first intron contribute to transcriptional
 RT control of the human alpha 1(I) collagen gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
 RN [13]
 RP SEQUENCE OF 1-44 FROM N.A.
 RX MEDLINE=88033098; PubMed=2822714;
 RA Rosouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
 RA de Wet W.;
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)
 RT collagen gene enhance transcription.";
 RL J. Biol. Chem. 262:15151-15157(1987).
 RN [14]
 RP SEQUENCE OF 33-52.
 RX PubMed=2318855;
 RA Wirtz M.K., Keene D.R., Hori H., Ghanville R.W., Steinmann B.,
 RA Rao V.H., Hollister D.W.;
 RT "In vivo and in vitro noncovalent association of excised alpha 1(I)
 RT amino-terminal propeptides with mutant alpha 2(I) collagen chains
 RT in native mutant collagen in a case of Ehlers-Danlos syndrome, type
 RT VII.";
 RL J. Biol. Chem. 265:6312-6317(1990).
 RN [15]
 RP SEQUENCE OF 156-183 FROM N.A.
 RX PubMed=2767050;
 RA Weil D., D'Alessio M., Ramirez F., de Wet W., Cole W.G., Chan D.,
 RA Bateman J.F.;
 RT "A base substitution in the exon of a collagen gene causes alternative
 RT splicing and generates a structurally abnormal polypeptide in a
 RT patient with Ehlers-Danlos syndrome type VII.";
 RL EMBO J. 8:1705-1710(1989).
 RN [16]
 RP SEQUENCE OF 175-187 AND 274-289.
 RX PubMed=2169412;
 RA Baetge B., Notbohm H., Diebold J., Lehmann H., Bodo M., Deutzmann R.,
 RA Muller P.K.;
 RT "A critical crosslink region in human-bone-derived collagen type I.
 RT Specific cleavage site at residue Leu95.";
 RL Eur. J. Biochem. 192:153-159(1990).
 RN [17]
 RP SEQUENCE OF 263-268.
 RX TISSUE=Skin;
 RA MEDLINE=71001508; PubMed=4319110;
 RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
 RT "A comparative study of glycopeptides derived from selected vertebrate
 RT collagens. A possible role of the carbohydate in fibril formation.";
 RL J. Biol. Chem. 245:5042-5048(1970).
 RN [18]
 RP SEQUENCE OF 281-302; 402-420; 823-842; 1026-1045 AND
 RP 1143-1162 FROM N.A.
 RX PubMed=2374517;
 RA Labhard M.E., Hollister D.W.;

RT "Segmental amplification of the entire helical and telopeptide regions
 RT of the cDNA for human alpha 1(I) collagen.";
 RL Matrix 10:124-130(1990).
 RN [19]
 RP SEQUENCE OF 425-490; 965-1024; 999-1039 AND 1453-1464 FROM N.A.
 RX MEDLINE=83064528; PubMed=6183642;
 RA Chu M.-L., Myers J.C., Bernard M.P., Ding J.-F., Ramirez F.;
 RT "Cloning and characterization of five overlapping cDNAs specific for
 RT the human pro alpha 1(I) collagen chain.";
 RL Nucleic Acids Res. 10:5925-5934(1982).
 RN [20]
 RP SEQUENCE OF 710-745 FROM N.A., AND VARIANT OI-II ARG-728.
 RX PubMed=2339700;
 RA Wallis G.A., Starman B.J., Zimm A.B., Byers P.H.;
 RT "Variable expression of osteogenesis imperfecta in a nuclear family is
 RT explained by somatic mosaicism for a lethal point mutation in the
 RT alpha 1(I) gene (COL1A1) of type I collagen in a parent.";
 RL Am. J. Hum. Genet. 46:1034-1040(1990).
 RN [21]
 RP SEQUENCE OF 746-781 FROM N.A., AND VARIANT OI-III SER-767.
 RX MEDLINE=95187161; PubMed=7881420;
 RA Forlino A., Zolezzi F., Valli M., Pignatti P.F., Cetta G.,
 RA Brunelli P.C., Mottes M.;
 RT "Severe (type III) osteogenesis imperfecta due to glycine
 RT substitutions in the central domain of the collagen triple helix.";
 RL Hum. Mol. Genet. 3:2201-2206(1994).
 RN [22]
 RP SEQUENCE OF 1187-1220 FROM N.A., AND VARIANT CYS-1195.
 RX PubMed=3170557;
 RA Cohn D.H., Apone S., Eyre D.R., Starman B.J., Andreassen P.,
 RA Charbonneau H., Nicholls A.C., Pope F.M., Byers P.H.;
 RT "Substitution of cysteine for glycine within the carboxyl-terminal
 RT telopeptide of the alpha 1 chain of type I collagen produces mild
 RT osteogenesis imperfecta.";
 RL J. Biol. Chem. 263:14605-14607(1988).
 RN [23]
 RP SEQUENCE OF 1440-1464 FROM N.A.
 RX MEDLINE=90110490; PubMed=2295701;
 RA Willing M.C., Cohn D.H., Byers P.H.;
 RT "Frameshift mutation near the 3' end of the COL1A1 gene of type I
 RT collagen predicts an elongated pro alpha 1(I) chain and results in
 RT osteogenesis imperfecta type I.";
 RL J. Clin. Invest. 85:282-290(1990).
 RN [24]
 RP SEQUENCE OF 1454-1464 FROM N.A.
 RX MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
 RA Maatta A., Bornstein P., Penttinen R.P.;
 RT "Highly conserved sequences in the 3'-untranslated region of the
 RT COL1A1 gene bind cell-specific nuclear proteins.";
 RN [25]
 RP Query Match 97.0%; Score 1149; DB 1; Length 1464;
 RP Best Local Similarity 93.6%; Pred. No 7,7e-44;
 RP Matches 206; Conservative 0; Mismatches 2; Indels 12; Gaps 1;
 QY 1 GPP-----GEPGPTGLPGPPGRRGGGRRGFGGADGAGVAGPKGPGAGSGPGA 48
 Db |||||
 Db 458 GPPGPAAGEGKRGARGEGPTGLPGPPGRRGGGRRGFGGADGAGVAGPKGPGAGSGPGA 517
 QY 49 GPKGSPCEAGRPGCEAGLPGCAKGLTGSPGSPGDKTGGPPGAGQDGRPPGPPGARGQA 108
 Db |||||
 Db 518 GPKGSPCEAGRPGCEAGLPGCAKGLTGSPGSPGDKTGGPPGAGQDGRPPGPPGARGQA 577
 QY 109 GVGWFFGPKGAAGEFGKAGERGVPPGPAVGAGKDGAGAGQPPGPAAGPAGERGQPGA 168
 Db 578 GVGWFFGPKGAAGEFGKAGERGVPPGPAVGAGKDGAGAGQPPGPAAGPAGERGQPGA 637
 QY 169 GSPGFQGLPGPAGPCEACKPGCEQGVPGDLGAPGSPGAG 208
 Db 638 GSPGFQGLPGPAGPCEACKPGCEQGVPGDLGAPGSPGAG 677

RESULT 5
 Q8N473

Qy	109	GVNGFFQGLPGKAAAGPGKAGRGVGGPPCAVGPAGKDGGAQAQGGPPGAGRGEGQGA	168
Db	578	GVNGFFQGLPGKAAAGPGKAGRGVGGPPCAVGPAGKDGGAQAQGGPPGAGRGEGQGA	637
Qy	169	GSPGFQGLPGPAGPPGAEAGKPGQGVGGVGDGAGPAGSPGAG	208
Db	638	GSPGFQGLPGPAGPPGAEAGKPGQGVGGVGDGAGPAGSPGAG	677
RESULT 6			
ID	Q63079	PRELIMINARY; PRT; 1453 AA.	
AC	Q63079;		
DT	01-NOV-1996	(TrEMBLrel. 01, Created)	
DT	01-JUN-1998	(TrEMBLrel. 06, Last sequence update)	
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)	
DE	Collagen alpha1 (Fragment).		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Sprague-Dawley; TISSUE=Bone, and Tooth;		
RX	MEDLINE=9913824; PubMed=1005941;		
RA	Brandsten C., Lundmark C., Christersson C., Hammarstrom L., Wurtz T.;		
RA	"Expression of collagen alpha1(I) mRNA variants during tooth and bone		
RL	formation in the rat.,"		
RL	J. Dent. Res. 78:11-19(1999).		
DR	EMBL; Z78279; CAB01633.1; ..		
DR	GO; GO:0005581; C:collagen; IEA.		
DR	GO; GO:0005737; C:cytoplasm; IEA.		
DR	GO; GO:0005201; F:extracellular matrix structural constituent; IEA.		
DR	GO; GO:0006817; P:phosphate transport; IEA.		
DR	InterPro; IPR008161; Clg helix.		
DR	InterPro; IPR008160; Collagen.		
DR	InterPro; IPR000885; Fib collagen_C.		
DR	InterPro; IPR009041; PMP_SGCI.		
DR	InterPro; IPR001007; VWF_C.		
DR	Pfam; PF01410; COLFI; 1.		
DR	Pfam; PF01391; Collagen; 18.		
DR	ProDom; PD000007; Clg helix; 3.		
DR	ProDom; PD002078; Fib collagen_C; 1.		
DR	SMART; SM00038; COLFI; 1.		
DR	SMART; SM00214; VWC; 1.		
DR	PROSITE; PS01208; VWFC 1; 1.		
DR	PROSITE; PS01084; VWFC_2; 1.		
KW	Collagen.		
FT	NON TER.		
FT	SEQUENCE 1 1		
SO	SEQUENCE 1453 AA; 137886 MW; E6896BDC19A41D8 CRC64;		
Query Match			
Best Local Similarity 94.8%; Score 1122; DB 2; Length 1453;			
Matches 201; Conservative 2; Mismatches 5; Indels 12; Gaps 1;			
Qy	1	GPP-----GSPGTGLPGPPGRRGGPGSGRFPAGDVAGVKGPGAGRGSPGA	48
Db	447	GPPGAGEGKRGARGGPGPSGLPGPPGRRGGPGSGRFPAGDVAGVKGPGSGSPGA	506
Qy	49	GPKGSPEAGRPGEAGLPGAKGLTGPSGSPGDKTGPAGODGPRGPPGPGARGQA	108
Db	507	GPKGSPEAGRPGEAGLPGAKGLTGPSGSPGDKTGPAGQDGRPPAGPPGARGQA	566
Qy	109	GVNGFFQGLPGPAGPPGAEAGKPGQGVGGVGDGAGPAGSPGAG	208
Db	567	GVNGFFQGLPGPAGPPGAEAGKPGQGVGGVGDGAGPAGSPGAG	626
Qy	169	GSPGFQGLPGPAGPPGAEAGKPGQGVGGVGDGAGPAGSPGAG	208
Db	627	GSPGFQGLPGPAGPPGAEAGKPGQGVGGVGDGAGPAGSPGAG	666

CALL_RAT	CALL_RAT	STANDARD;	PRT;	671 AA.
AC	P02454;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	25-OCT-2004 (Rel. 45, Last annotation update)			
DE	Collagen alpha 1(I) chain (Fragments).			
GN	Name=Collal;			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
NCBI_TaxID=10116;				
[1]				
RP	SEQUENCE OF 1-19.			
RP	MEDLINE=69155173; PubMed=5777344;			
RA	Bornstein P.;			
RT	"Comparative sequence studies of rat skin and tendon collagen. II. The			
RT	absence of a short sequence at the amino terminus of the skin alpha-			
RT	chain.";			
RL	Biochemistry 8:63-71(1969).			
RL	[2]			
RP	SEQUENCE OF 5-19.			
RP	MEDLINE=67162268; PubMed=5337886;			
RA	Kang A.H., Bornstein P., Piez K.A.;			
RT	"The amino acid sequence of peptides from the cross-linking region of			
RT	rat skin collagen.";			
RL	Biochemistry 6:788-795(1967).			
RL	[3]			
RP	SEQUENCE OF 20-55.			
RP	MEDLINE=67165368; PubMed=4290711;			
RA	Bornstein P.;			
RT	"The incomplete hydroxylation of individual prolyl residues in			
RT	collagen.";			
RL	J. Biol. Chem. 242:2572-2574(1967).			
RL	[4]			
RP	SEQUENCE OF 56-102.			
RP	MEDLINE=71263178; PubMed=4327399;			
RA	Butler W.T., Ponds S.L.;			
RT	"Chemical studies on the cyanogen bromide peptides of rat skin			
RT	collagen. Amino acid sequence of alpha 1-CB4.";			
RL	Biochemistry 10:2076-2081(1971).			
RL	[5]			
RP	SEQUENCE OF 103-139.			
RP	MEDLINE=70085124; PubMed=5411206;			
RA	Butler W.T.;			
RT	"Chemical studies on the cyanogen bromide peptides of rat skin			
RT	collagen. The covalent structure of alpha 1-CB5, the major hexose-			
RT	containing cyanogen bromide peptide of alpha 1.";			
RL	Biochemistry 9:44-50(1970).			
RL	[6]			
RP	SEQUENCE OF 140-238.			
RP	MEDLINE=72136131; PubMed=4335087;			
RA	Balian G., Click E.M., Bornstein P.;			
RT	"Structure of rat skin collagen alpha 1-CB8. Amino acid sequence of			
RT	the hydroxylamine-produced fragment HA1.";			
RL	Biochemistry 10:4470-4478(1971).			
RL	[7]			
RP	SEQUENCE OF 239-418.			
RP	MEDLINE=73006942; PubMed=4342027;			
RA	Balian G., Click E.M., Hermanson M.A., Bornstein P.;			
RT	"Structure of rat skin collagen alpha 1-CB8. Amino acid sequence of			
RT	the hydroxylamine-produced fragment HA2.";			
RL	Biochemistry 11:3798-3806(1972).			
RL	[8]			
RP	SEQUENCE OF 419-567.			
RP	MEDLINE=74271984; PubMed=4366532;			
RA	Butler W.T., Underwood S.P., Finch J.E. Jr.;			
RT	"Chemical studies on the cyanogen bromide peptides of rat skin			
RT	collagen. Amino acid sequence of alpha 1-CB3.";			
RL	Biochemistry 13:2946-2953(1974).			
RL	[9]			
RP	SEQUENCE OF 568-651.			
RP	MEDLINE=74011954; PubMed=4126850;			
RP	[10]			

Qy	1	GPP-----GEPGPTGLPQPGERGGPSRFPFGADGVAGPKGPAGERSGPGPA	48
Df	447	: : : : : : : : : :	506
Qy	49	GPKGSPGEARNGEAGLTPKAKGLTGSFSGSDPDGKTGPPGAGDGRPGPGPPGARQA	108
Df	507	GPKSFGSEAGRPEAGLTGAAGLTSFSGSPDGKTGPPGAGDGRPGPGPPGARQA	566
Qy	109	GVWGPFPGKAAGEBKAGRGVPGAVGVPAGCKGCAQCPGAPGAGRGQGPA	168
Df	567	GVWGFPGKGTAGEPCKAGERKLFGPPGAVGVPAGCKGCAQCAPGAPGAGRGQGPA	626
Qy	169	GSFGFQGLPGAGPPGEAGKPEGVGFDLGAPGPSGPA	208
Df	627	GSFGFQGLPGAGPPGEAGKPEGVGFDLGAPGPSGARG	666

RESULT 10

```

Q810J9 PRELIMINARY; PRT; 1453 AA.
ID Q810J9 AC
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DE 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Procollagen, type I, alpha 1.
GN Name=Collai;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxId=10090;
[1] SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RX MEDLINE=223388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
RN [2]

SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC050014; AAH50014.1; -.
DR MGD; MG1:88467; Collai.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPRO08161; Clg helix.
DR InterPro; IPRO08160; Collagen.
DR InterPro; IPRO00885; Fib collagen_C.
DR InterPro; IPRO09041; PME_SGCI-.
DR InterPro; IPRO01007; VWF-C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; WVC; 1.
DR ProbDom; PD000007; Clg helix; 3.
DR ProbDom; PD002078; Fib collagen_C; 1.

```


Db 444 GPPGSGEEGKRGSRGBFPAGPPGPPAGERGCGPSRGPFGSDGASGPKAPAGERSVGPA 503

Qy 49 GPKGSGEAGRGPGEAGLPGAKGLTGS PGSPGDKTGTPPGPAGQDGRPGPPGPGARGOA 108
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 Db 504 GPKGSTGESGEPGPGPLPGAKGLTGS PGSPGDKTGTPAGAAQDGHGPPGPGSGARGQS 563
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 Qy 109 GVMGFPGPKGAAGPGKAGRGVFPGPAGVAGPAGKDGAGAGQGPFGPAGPAGERGEQGPA 168
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 Db 564 GVMGFPGPKGAAGBFGKSGRGVAGPPGATGAPGKDGEGAGAQGPFGPSGSGERGEGQA 623
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 Qy 169 GSPGFGQLPGPAGPGGAGKDGEGVPGDILGAPGSPGAG 208
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 Db 624 GSPGFGQLPGSPGAGAGKPGEGQAGPGDAGGPGSPGRG 663
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 RESULT 13
 O93251 PRELIMINARY; PRT; 1445 AA.
 ID O93251 AC O93251;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Alpha 1 type I collagen.
 GN Name=alpha 1 type I collagen;
 OS Rana catesbeiana (Bull frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Rana.
 OX NCBI_TaxID=8400;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99294154; PubMed=10367734; DOI=10.1016/S0945-053X(99)00005-0;
 RA Aashina K., Utoh R., Obara M., Yoshizato K.;
 RT "Cell-type specific and thyroid hormone-dependent expression of genes
 of $\alpha 1(I)$ and $\alpha 2(I)$ collagen in intestine during
 RT amphibian metamorphosis.";
 RL Matrix Biol. 18:89-103(1999).
 DR EMBL; AB015440; BAA29028.1; -
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005737; C:cyttoplasm; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR00885; Fib collagen_C.
 DR InterPro; IPR009041; PMP_SGCI.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; Clg helix; 4.
 DR ProDom; PD002078; Fib collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWFC_1; 1.
 DR PROSITE; PS00184; VWFC_2; 1.
 KW Collagen.
 SQ SEQUENCE 1445 AA; 137251 MW; F59BB550C9873F04 CRC64;
 Query Match 82.9%; Score 981; DB 2; Length 1445;
 Best Local Similarity 82.0%; Pred. No. 2.1e-36;
 Matches 178; Conservative 5; Mismatches 22; Indels 12; Gaps 2
 Qy 1 GPP-----CEPPTGLPGPPGRRGGPSRGPFGADGVAGPKPAGERGSGPGPAGPK 51
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 Db 446 GPGGEGKRSRGPPGSPGPPGAPGARGAPGSRGPPGADGAGGKPPGRRGPGVSGAPK 505
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 Qy 52 GSPGEGRGPGEAGLPGAKGLTGS PGSPGDKTGTPPGAGQDGRPGPPGPGARGOAGVM 111
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 Db 506 GSPGSGRGPGEPLPGAKGLTGS PGSPGDKTGTPAGAPGDGRPGPPGARGOQSGVM 565
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 Qy 112 GFGPGKAAAGEPKAGBRGVPPGPAVGPAKDGAGAGQGPFGPAGPAGERGEQGPA 171
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :
 Db 566 GFGPGKAAAGEPKPGBRGVAGPGVAGAPGKDGAGAGQGPFGPAGPAGERGEQGPA 625
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| :

QY 172 GFQGLPGPAGPGAGKPGEGVGDLCAGPSPGAG 208
|||||
Db 626 GFQGLPGSPGAGSGKPGEGVGDV---GPSGPAG 659

RESULT 14

Q640B2 PRELIMINARY; PRT; 1449 AA.
ID Q640B2
AC Q640B2
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole body;
RX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Shenmen C.M., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Schaefer C.F., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Hsieh F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole body;
RX Klein S., Gerhard D.S.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
EMBL: BC082718; AAH82718.1;
KW Hypothetical protein.
SQ SEQUENCE 1449 AA; 137539 MW; 456639B1687A3B4B CRC64;

Query Match 82.3%; Score 974.5; DB 2; Length 1449;
Best Local Similarity 80.0%; Pred. No. 4.1e-36;
Matches 176; Conservative 9; Mismatches 20; Indels 15; Gaps 2;

QY 1 GPP-----GEPGTGLCPGGERGPGSRGPPGAGVAGKPGAGRGSPGPA 48
Db 444 GPPPSGEGKRGSRGEPGPPGAGPAGCAPGSRGPPGSDGASGKPGGPGVCPA 503

QY 49 GPKSGPGEAGPGEAGLPGKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQA 108
Db 504 GAKSGPGEAGPGEAGLPGKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQS 563

QY 109 GVMGPPGKGAAGEPGKAGRGVPGPPGAVGAPGAGKAGQPPGPPGAGRGEGQGA 168
Db 564 GVMGPPGKGAAGEPGKAGRGVPGPPGAVGAPGAGLPGKDGAGAGQPPGPPGAGRGEGQGA 623

QY 169 GSPGFGQLGPPAGPGEAGKPGEGVGDLCAGPSPGAG 208
Db 624 GPGFGQLGPPAGPGEAGKPGEGVGDV---GPSGPAG 660

RESULT 15

Query Match 81.3%; Score 962.5; DB 2; Length 1449;

Q802B5 PRELIMINARY; PRT; 1449 AA.
ID Q802B5
AC Q802B5
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Collal-prov protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RA "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative";
RL Dev. Dyn. 225:384-391 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
EMBL: BC049829; AAH49829.1;
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005737; C:cyttoplasm; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0006817; P:phosphate transport; IEA.
DR InterPro: IPR008161; Clg helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR009041; PMP_SGCI.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 18.
DR Pfam: PF00093; VWC; 1.
DR ProDom: PD000007; Clg helix; 4.
DR ProDom: PD002078; Fib collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
DR SMART: SM00214; VWC; 1.
DR PROSITE: PS01208; VWF_C_1; UNKNOWN_1.
DR PROSITE: PS0184; VWF_2; 1.
KW Collagen.
SQ SEQUENCE 1449 AA; 137464 MW; 4B32E20BECCE978C CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:15:59 ; Search time 100.815 Seconds
(without alignments)
2367.017 Million cell updates/sec

Title: US-10-658-989A-2
Perfect score: 3488
Sequence: 1 GPPGEPGPTGLPGRPGRGG.....GEQVPGDLGAPGSPGAGG 617

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3488	100.0	617	8 ADM48391	Adm48391 Human rec
2	3488	100.0	617	8 ADQ26217	Adq26217 Human gel
3	3482	99.8	821	8 ADM48392	Adm48392 Human rec
4	2237.5	64.1	1057	3 AAY84544	Aay84544 A human c
5	2237.5	64.1	1058	3 AAY84403	Aay84403 Amino aci
6	2237.5	64.1	1107	3 AAY84540	Aay84540 Amino aci
7	2237.5	64.1	1171	3 AAY84538	Aay84538 A chimeri
8	2237.5	64.1	1464	2 AAW68485	Aaw68485 Human rec
9	2237.5	64.1	1464	7 ADD45059	Add45059 Human Pro
10	2237.5	64.1	1464	7 ADD45055	Add45055 Human Pro
11	2237.5	64.1	1464	7 ADD45051	Add45051 Human Pro
12	2237.5	64.1	1464	8 ADRI16800	Adri16800 Human col
13	2237.5	64.1	1464	8 ADRI16425	Adri16425 Human col
14	2237.5	64.1	1464	8 ADRI16425	Adri16425 Human col
15	2234	64.0	1449	4 AAE02535	Aae02535 Porcine a
16	2233.5	64.0	1161	7 ADE87050	Ade87050 Human pan
17	2233.5	64.0	1461	5 ABG93947	Abg93947 Human pol
18	2233.5	64.0	1464	4 AAU14136	Aau14136 Human nov
19	2233.5	64.0	1536	7 ADE87051	Ade87051 Human pan
20	2233.5	64.0	1057	3 AAY84541	Aay84541 Amino aci
21	2229.5	63.9	1388	3 AAY84539	Aay84539 Amino aci
22	2228.5	63.9	1341	2 AAR71701	Aar71701 Collagen
23	2228.5	63.9	1341	3 AAY96122	Aay96122 Collagen
24	2228.5	63.9	1341	5 AAE16475	Aae16475 Human col
25	2228.5	63.9	1341	5 ABB80733	Abb80733 Collagen

26	2228.5	63.9	1341	5 ABB09625	Abb09625 Amino aci
27	2228.5	63.9	1341	7 ADF13075	Adf13075 Human col
28	2227.5	63.9	1464	4 AAB82454	Aab82454 Human pro
29	2227.5	63.9	1464	5 ABB90764	Abb90764 Human tum
30	2227.5	63.9	1464	5 ABP68610	Abp68610 Human pan
31	2227.5	63.9	1464	6 ABU54471	Abu54471 Human tum
32	2227.5	63.9	1464	6 ABR47417	Abr47417 Breast ca
33	2227.5	63.9	1464	6 ABR92064	Abr92064 Human cer
34	2227.5	63.9	1464	7 ADD14142	Add14142 Human src
35	2227.5	63.9	1464	7 ADP65246	Adp65246 Human alp
36	2227.5	63.9	1464	8 ADQ19470	Adq19470 Human sof
37	2227.5	63.9	1464	8 ADQ29653	Adq29653 Human col
38	2225.5	63.8	1463	4 AAE02532	Aae02532 Bovine al
39	2224.5	63.8	1464	7 ADE87048	Ade87048 Human pan
40	2222.5	63.7	1107	2 AAR89472	Aar89472 Collagen/
41	2222.5	63.7	1169	2 AAR89469	Aar89469 Collagen/
42	2222.5	63.7	1169	3 AAY84537	Aay84537 Amino aci
43	2222.5	63.7	1171	2 AAR89470	Aar89470 Collagen/
44	2222.5	63.7	1388	2 AAR89471	Aar89471 Collagen/
45	2222	63.7	1211	7 ADE87057	Ade87057 Human pan

ALIGNMENTS

RESULT 1

ADM48391
ID ADM48391 standard; protein; 617 AA.

AC ADM48391;

XX 03-JUN-2004 (first entry)

DE Human recombinant gelatin-like polypeptide Hu-3.

KW Plasma substitute; Gelatin-like protein; plasma expander; human.

XX Homo sapiens.

XX EPI398324-A1.

PD 17-MAR-2004.

XX 11-SEP-2002; 2002EP-00078745.

XX 11-SEP-2002; 2002EP-00078745.

XX (FUJF) FUJI PHOTO FILM BV.

XX Bouwstra JB, Toda Y;

XX WPI; 2004-229415/22.

PT Composition useful as substitute for plasma, comprises solution of saline and recombinant gelatin-like protein having colloid osmotic function.

XX Example 1; SEQ ID NO 2; 31pp; English.

CC The present sequence is the protein sequence of recombinant gelatin-like protein Hu-3. This is a trimer of human recombinant gelatin-like protein Hu-1 ADM48390. A claimed composition suitable as a substitute for plasma comprises a solution of saline and a protein having a colloid osmotic function. The protein is a recombinant gelatin-like protein with a molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric point of less than 8. It is especially Hu-1 or Hu-deam ADM48393. The recombinant gelatin-like protein, or a dimer, trimer or tetramer of the protein, is useful as a plasma expander that has a lower clearance rate from blood circulation, provides better and predictable regulation of clearance rate and which is less susceptible to proteolytic degradation than presently used gelatin derivatives. Recombinant gelatin-like proteins that are in essence free of hydroxyproline do not give rise to an immunological reaction with blood samples containing IgE antibodies. The gelatin-like proteins can be covalently attached to pharmaceutically

CC active compounds. After administration, the coupled medicament will not
 CC diffuse from the circulating blood into the interstitium. Clearance by
 CC liver and kidney will be kept to a minimum, ensuring a more constant
 CC plasma level of the medicament. Suitable medicaments include those
 CC involved in intervening blood clotting, vasodilation, function of
 CC erythrocytes, thrombocytes and leukocytes, immune responses, and blood
 CC levels of messenger molecules such as hormones.
 XX
 SQ Sequence 617 AA;

Query Match 100.0%; Score 3488; DB 8; Length 617;
 Best Local Similarity 100.0%; Pred. No. 1.8e-207;
 Matches 617; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPPGPGTGLPDPGPGGSGRFPAGDVGAGKPGAGRGSPGAGKSGPGEAGRP 60
 Db 1 GPPGPGTGLPDPGPGGSGRFPAGDVGAGKPGAGRGSPGAGKSGPGEAGRP 60
 QY 61 GEAGLPGAKGLTSGSPGDKTGPAGDGRGPPGPPGAGQAGVWGFPGKGA 120
 Db 61 GEAGLPGAKGLTSGSPGDKTGPAGDGRGPPGPPGAGQAGVWGFPGKGA 120
 QY 121 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 180
 Db 121 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 180
 QY 181 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 240
 Db 181 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 240
 QY 241 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 300
 Db 241 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 300
 QY 301 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 360
 Db 301 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 360
 QY 361 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 420
 Db 361 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 420
 QY 421 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 480
 Db 421 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 480
 QY 481 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 540
 Db 481 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 540
 QY 541 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 600
 Db 541 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 600
 QY 601 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 617
 Db 601 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 617

RESULT 2
 ADQ26217
 ID ADQ26217 standard; protein; 617 AA.
 XX
 AC ADQ26217;
 XX
 DT 23-SEP-2004 (first entry)
 XX
 DE Human gelatine-like polypeptide Hu-3.
 KW Human; gelatine-like protein; Hu-3; microcarrier; cell culture.
 XX Homo sapiens.
 OS
 XX

PN WO2004056976-A2.
 XX 08-JUL-2004.
 XX 23-DEC-2003; 2003WO-NL000922.
 XX 23-DEC-2002; 2002EP-00080539.
 XX (FUJF) FUJI PHOTO FILM BV.
 XX Bouwstra JB, Van Es AJJ, Toda Y;
 XX WPI; 2004-507711/48.
 XX
 XX Preparing cell culture support useful for culturing anchorage dependent
 PT cells, involves coating microcarrier bead with gelatine or gelatine-like
 PT protein having specific molecular weight.
 XX
 XX Example 1; SEQ ID NO 1; 19pp; English.
 XX
 XX The present sequence is the protein sequence of human recombinant gelatin
 CC -like polypeptide Hu-3, which has a molecular weight of approximately 54
 CC kDa and which contains multiple Gly-Xaa-Yaa triplets. In an example from
 CC the invention, recombinant Hu-3 was immobilised on polystyrene beads
 CC using a heterobifunctional crosslinking agent, and used in a spinner
 CC flask cell culture. A claimed process for the preparation of a cell
 CC culture support comprises the step of coating a microcarrier bead with
 CC gelatine or gelatine-like protein having a molecular weight of about 40-
 CC 200 kDa, and optionally further comprising the step of immobilising the
 CC gelatine or gelatine-like protein on the microcarrier. In this process,
 CC more than 75%, preferably more than 85% and more preferably more than 95%
 CC of the gelatine or gelatine-like protein has the same molecular weight.
 CC The gelatine or gelatine-like protein is recombinantly produced to obtain
 CC a material of uniform molecular weight and to reduce the risk of
 CC contamination with prions. A claimed cell support consists of microbeads
 CC of 50-500 µm size coated with a gelatine-like protein consisting of at
 CC least 95% Gly-Xaa-Yaa triplets and containing at least 15% proline
 CC residues and less than 5% of hydroxyproline residues, with a molecular
 CC weight distribution showing a maximum between 40 and 200 kDa, at least
 CC 75% of the protein molecules having a molecular weight within 2% of the
 CC maximum. Large-scale production of expressed products can be accomplished
 CC with gelatine-coated microcarriers.
 XX
 SQ Sequence 617 AA;

Query Match 100.0%; Score 3488; DB 8; Length 617;
 Best Local Similarity 100.0%; Pred. No. 1.8e-207;
 Matches 617; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPPGPGTGLPDPGPGGSGRFPAGDVGAGKPGAGRGSPGAGKSGPGEAGRP 60
 Db 1 GPPGPGTGLPDPGPGGSGRFPAGDVGAGKPGAGRGSPGAGKSGPGEAGRP 60
 QY 61 GEAGLPGAKGLTSGSPGDKTGPAGDGRGPPGPPGAGQAGVWGFPGKGA 120
 Db 61 GEAGLPGAKGLTSGSPGDKTGPAGDGRGPPGPPGAGQAGVWGFPGKGA 120
 QY 121 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 180
 Db 121 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 180
 QY 181 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 240
 Db 181 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 240
 QY 241 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 300
 Db 241 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 300
 QY 301 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 360
 Db 301 GPPGAGRGVPPGPGVAGPKDGEAGAGQPPGPPAGRGSGPAGSGPFGQLP 360

QY 361 GPAGERGQPSGPGFQGLPGPAGPCEAGKPGCEQGVGCDLGAAPSGPAGPCEPTGLP 420
Db |||||
QY 361 GPAGERGQPSGPGFQGLPGPAGPCEAGKPGCEQGVGCDLGAAPSGPAGPCEPTGLP 420
Db |||||
QY 421 GPPGERGPGSRGPGADGVAGPKGPGAGERSGPGAGPKGSPGAGLPGAKGLT 480
Db |||||
QY 481 GSPGSPGPDGKTGTPPGAGQDGRPGPPGARGQAGVMGPPGKGAAGBPKAGRGV 540
Db |||||
QY 541 GPPGAVGPGAGKDGAGAGQPPGPPGAGPAGERGQPSGPGFQGLPGPAGPCEAGKPGCEQ 600
Db |||||
QY 601 GVPDGLGAPGSPGAGG 617
Db |||||
RESULT 3
ID ADM48392 standard; protein; 821 AA.
XX
AC ADM48392;
XX
DT 03-JUN-2004 (first entry)
XX
DE Human recombinant gelatin-like polypeptide Hu-4.
XX
KW Plasma substitute; Gelatin-like protein; plasma expander; human.
XX
OS Homo sapiens.
XX
PN EP1398324-A1.
XX
PD 17-MAR-2004.
XX
PF 11-SEP-2002; 2002EP-00078745.
XX
PR 11-SEP-2002; 2002EP-00078745.
XX
PA (FUJF) FUJI PHOTO FILM BV.
XX
PI Bouwstra JB, Toda Y;
XX
DR WPT; 2004-229415/22.
XX
PT Composition useful as substitute for plasma, comprises solution of saline
XX and recombinant gelatin-like protein having colloid osmotic function.
XX
PS Example 1; SEQ ID NO 3; 31pp; English.
XX
CC The present sequence is the protein sequence of recombinant gelatin-like
CC protein Hu-3. This is a tetramer of human recombinant gelatin-like
CC protein Hu-1 ADM48390. A claimed composition suitable as a substitute for
CC plasma comprises a solution of saline and a protein having a colloid
CC osmotic function. The protein is a recombinant gelatin-like protein with
CC a molecular weight of at least 10 kDa and at most 50 kDa and an
CC isoelectric point of less than 8. It is especially Hu-1 or Hu-deam
CC ADM48393. The recombinant gelatin-like protein, or a dimer, trimer or
CC tetramer of the protein, is useful as a plasma expander that has a lower
CC clearance rate from blood circulation, provides better and predictable
CC regulation of clearance rate and which is less susceptible to proteolytic
CC degradation than presently used gelatin derivatives. Recombinant gelatin-
CC like proteins that are in essence free of hydroxyproline do not give rise
CC to an immunological reaction with blood samples containing IgE
CC antibodies. The gelatin-like proteins can be covalently attached to
CC pharmaceutically active compounds. After administration, the coupled
CC medicament will not diffuse from the circulating blood into the
CC interstitium. Clearance by liver and kidney will be kept to a minimum,
CC ensuring a more constant plasma level of the medicament. Suitable

CC medicaments include those involved in intervening blood clotting,
CC vasodilation, function of erythrocytes, thrombocytes and leukocytes,
CC immune responses, and blood levels of messenger molecules such as
CC hormones.
XX
SQ Sequence 821 AA;
Query Match 99.8%; Score 3482; DB 8; Length 821;
Best Local Similarity 100.0%; Pred. No. 5.2e-207;
Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GPPGEPDPTGLPFPGERGPGSRGPGADGVAGPKGPGAGERSGPGAGPKGSPGAGRP 60
Db 1 GPPGEPDPTGLPFPGERGPGSRGPGADGVAGPKGPGAGERSGPGAGPKGSPGAGRP 60
QY 61 GEAGLPGAKGLTSGSPGPDGKTGTPPGAGQDGRPGPPGARGQAGVMGPPGPKGAA 120
Db 61 GEAGLPGAKGLTSGSPGPDGKTGTPPGAGQDGRPGPPGARGQAGVMGPPGPKGAA 120
QY 121 GEPGKAGRGVPGPPGAVGPGAGKDGAGAGQPPGAGPAGERGQPSGPGFQGLPGPA 180
Db 121 GEPGKAGRGVPGPPGAVGPGAGKDGAGAGQPPGAGPAGERGQPSGPGFQGLPGPA 180
QY 181 GPPGKAGRGVPGPPGAVGPGAGKDGAGAGQPPGAGPAGERGQPSGPGFQGLPGPA 240
Db 181 GPPGKAGRGVPGPPGAVGPGAGKDGAGAGQPPGAGPAGERGQPSGPGFQGLPGPA 240
QY 241 GPAGERSGPGAGPKGSPGAGRPGEAGLPGAKGLTSGSPGPDGKTGTPPGAGQDGRP 300
Db 241 GPAGERSGPGAGPKGSPGAGRPGEAGLPGAKGLTSGSPGPDGKTGTPPGAGQDGRP 300
QY 301 GPPGPPGARGQAGVMGPPGPKGAAAGBPKAGRGVPGPPGARGQAGVMGPPGAGQPPGPA 360
Db 301 GPPGPPGARGQAGVMGPPGPKGAAAGBPKAGRGVPGPPGARGQAGVMGPPGAGQPPGPA 360
QY 361 GPAGERGQPSGPGFQGLPGPAGPCEAGKPGCEQGVGCDLGAAPSGPAGPCEPTGLP 420
Db 361 GPAGERGQPSGPGFQGLPGPAGPCEAGKPGCEQGVGCDLGAAPSGPAGPCEPTGLP 420
QY 421 GPPGERGPGSRGPPGADGVAGPKGPGAGERSGPGAGPKGSPGAGLPGAKGLT 480
Db 421 GPPGERGPGSRGPPGADGVAGPKGPGAGERSGPGAGPKGSPGAGLPGAKGLT 480
QY 481 GSPGSPGPDGKTGTPPGAGQDGRPGPPGARGQAGVMGPPGKGAAGBPKAGRGV 540
Db 481 GSPGSPGPDGKTGTPPGAGQDGRPGPPGARGQAGVMGPPGKGAAGBPKAGRGV 540
QY 541 GPPGAVGPGAGKDGAGAGQPPGPPGAGPAGERGQPSGPGFQGLPGPAGPCEAGKPGCEQ 600
Db 541 GPPGAVGPGAGKDGAGAGQPPGPPGAGPAGERGQPSGPGFQGLPGPAGPCEAGKPGCEQ 600
QY 601 GVPDGLGAPGSPGAG 616
Db 601 GVPDGLGAPGSPGAG 616
RESULT 4
ID AAY84544 standard; protein; 1057 AA.
XX
AC AAY84544;
XX
DT 25-JUL-2000 (first entry)
XX
DE A human collagen 1 (alpha1) protein helical region.
XX
KW Extracellular matrix protein; self aggregation; hydroxylated proline;
XX trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
XX collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX
OS Homo sapiens.
XX
PN EP992586-A2.

XX PD 12-APR-2000.
 XX PF 07-OCT-1999; 99EP-00119184.
 XX PR 09-OCT-1998; 98US-00169768.
 XX PA (USSU) US SURGICAL CORP.
 XX PI Gruskin EA, Buechter DD, Zhang G, Connolly K;
 XX DR WPI; 2000-259138/23.
 XX DR N-PSDB; AA12503.
 XX PT Production of extracellular matrix proteins containing 4-trans-
 XX PT hydroxyproline results in native self aggregating proteins, useful on
 XX PT medical implants.
 XX PS Example 10; Fig 39A-E; 260pp; English.
 XX CC The specification describes a method for producing an extracellular
 XX CC matrix protein or its fragment. The extracellular matrix protein is
 XX CC capable of self aggregating in a cell which does not ordinarily
 XX CC hydroxylated prolines. The method comprises optimising a nucleic acid
 XX CC sequence for expression in the cell by substitution of codons preferred
 XX CC by that cell for naturally occurring codons not preferred by the cell;
 XX CC incorporating the nucleic acid sequence into the cell; and contacting the
 XX CC cell with a hypertonic growth medium containing at least one amino acid,
 XX CC selected from the group consisting of trans-4-hydroxyproline and 3-
 XX CC hydroxyproline to allow at least one of the amino acids to be assimilated
 XX CC into the cell and incorporated into the extracellular matrix protein. The
 XX CC method may be used to make host cells assimilate and incorporate trans-4-
 XX CC hydroxyproline into proteins. This is especially useful in the
 XX CC recombinant production of proteins such as collagen, fibrinogen and
 XX CC fibronectin whose ability to self aggregate and produce functional
 XX CC proteins depends on the post translational hydroxylation of proline. The
 XX CC method is also useful in studying the structure and function of
 XX CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 XX CC present sequence represents human collagen 1 (alpha1) helical region,
 XX CC which may be produced using the method of the invention
 XX SQ Sequence 1057 AA;
 Query Match 64.1%; Score 2237.5; DB 3; Length 1057;
 Best Local Similarity 60.2%; Pred. No. 3.2e-130;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;
 QY 1 GPP-----GEPGTLPGPPGRRGGRGFRGADGAGKPGAGERSGPA 48
 DB 297 GPPGAGEGKRGAGEGPGTGLPGPPGRRGGRGFRGADGAGKPGAGERSGPA 356
 QY 49 GPKGSPGEAGRPGCEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRFPDPPGARGQA 108
 DB 357 GPKGSPGEAGRPGCEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRFPDPPGARGQA 416
 QY 109 GVMGPPGPKGAAGEPGKAGRGVPPGAGVAGPKGDKGAGAGPPGPPGAGERGEQGPA 168
 DB 417 GVMGPPGPKGAAGEPGKAGRGVPPGAGVAGPKGDKGAGAGPPGPPGAGERGEQGPA 476
 QY 169 GSPFGQGLPGPAGPGEAGKPGEQGVGDLGAPGSPGAGE-----PQP----- 212
 DB 477 GSPFGQGLPGPAGPGEAGKPGEQGVGDLGAPGSPGAGE-----PQP----- 536
 QY 213 -----TGLPCCGRRGGRGSRGPPGADGAGVAGPKGAGERSGPPGAGRGP 258
 DB 537 GAGAPGNDGAKGDAGAPGAGQSGAPLQGMPPGRRGAAGLPGPKGDRGAGPKGAGDGP 596
 QY 259 GE-----AGRPGCEAGLPGAK-----GLTSGSPGPDGKTGPPGAGQDGRP 300
 DB 597 GKDVGRLGTGPTGPPGAGAPGDKGESGSPGAGTARGAPDGRGEPGPPGAGFAGPP 656
 QY 301 GPPGPPGARGQ-----AGVMGPPGPKGAAGEPGKAGRGVPPGPA----- 341

DB 657 GADGQPGKGEPCDAGAKGDAGPPGAGPAGPPGPIGNVAGPAKAGSAGPAGATGPP 716
 QY 342 -----VGPAGKDGGEAGAQGPPGAGP-----AGERGQGPAGSPGQGLPGPAGPPGAGKRP 393
 DB 717 GAAGRVPGPSPGNAGPPGPPGAGKEGKGRGETGAGREGEVGPFPGPPGAGKESGP 776
 QY 394 GEQGVPGDLGAPGSPGAGEPGTGLPGPPGRRGGRGFRGADGAGVAGPKGAPGERGSP 453
 DB 777 GADGAPAGPPTPGPQGIAGQGVVGLPGQRRGRRFPGLPGSGEFGKQGPSGASGERGPP 836
 QY 454 GPAGP---KGSFGEAGRPGCEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRFPGPPGPP 510
 DB 837 GVMGPPGLAGPPGSGREGAPGAEGRGPRDGSFGAKGDRGETGAGPAGPAGPAGP 896
 QY 511 GARGQAGVMGRRGPKGAGRGPKAGRGVPPGPPGAGVAGPKGDKGE-----AGAQ 558
 DB 897 GPAGKSGDRGETGTPAGPAGPVGPPGAGPAGPQGRGDKGETGQCGDRGKIGHRGFSGLQ 956
 QY 559 GPPGP-----AGPAGERGEQGPAGSP---GFGGLPGPAGPGEAGKPGEQGVP 603
 DB 957 GPPGPPGSGRQSGSGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1016
 QY 604 GDLGAPGSPGAG 616
 DB 1017 GPPGPPGPPGPPG 1029
 RESULT 5
 ID AAY84403 standard; protein; 1058 AA.
 AC AAY84403;
 DT 12-JUL-2000 (first entry)
 XX Amino acid sequence of human type 1 (alpha1) collagen polypeptide.
 DE Alpha1 collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;
 KW collagen; mussel adhesive protein; bioadhesive.
 OS Homo sapiens.
 XX WO200014201-A1.
 XX 16-MAR-2000.
 PF 07-SEP-1999; 99WO-US020462.
 PR 09-SEP-1998; 98US-0099852P.
 XX (USSU) US SURGICAL CORP.
 XX (PAOL/) PAOLELLA D N.
 XX (GRUS/) GRUSKIN E A.
 XX (BUEC/) BUECHTER D D.
 XX Paolella DN, Gruskin EA, Buechter DD;
 XX WPI; 2000-271051/23.
 DR N-PSDB; AAZ99843.
 PT Incorporating non-natural amino acid into polypeptide, useful e.g. for
 PT production of bioadhesives, by epoxidation or substitution of
 PT dehydroproline residues.
 XX Disclosure; Fig 6; 66pp; English.
 CC The present sequence represents a human type 1 (alpha1) collagen protein.
 CC Peptides derived from the protein were used to demonstrate incorporation
 CC of 3,4-dehydro-L-proline into the peptide, using the method of the
 CC invention. The specification describes a method for the incorporation of
 CC non-natural amino acid into a polypeptide. The method comprises reacting
 CC at least one 3,4-dehydroproline residue in the polypeptide with an
 CC epoxidation reagent from a polypeptide containing at least one 3,4-

CC epoxypyrrolone residue. The method is used for studying the effects of non-
 CC natural amino acids on structure and function of polypeptides. The method
 CC is also useful for commercial production of collagen or mussel adhesive
 CC proteins (which are useful as bioadhesives), and for incorporating a wide
 CC variety of groups, including therapeutic ligands and biological probes,
 CC into polypeptides
 XX
 SQ Sequence 1058 AA;

Query Match 64.1%; Score 2237.5; DB 3; Length 1058;
 Best Local Similarity 60.2%; Pred. No. 3.2e-130;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPCCGCGGSGRGGPGADGAGVAGKPGAGRGSGPGA 48
 DB 298 GPPGAGBEGKRGARGEPGPTGLPCCGCGGSGRGGPGADGAGVAGKPGAGRGSGPGA 357
 QY 49 GPKSGPGEAGRGAGLPGAKGLTSGSGSPGDKTGGPGAGODGRPPGPPGARGQA 108
 DB 358 GPKSGPGEAGRGAGLPGAKGLTSGSGSPGDKTGGPGAGODGRPPGPPGARGQA 417
 QY 109 GVMGFPKGKGAAGFGKAGRGVPGPGCAVGPAGKDGAGAGQGGPPGAPGAGERGEGQGA 168
 DB 418 GVMGFPKGKGAAGFGKAGRGVPGPGCAVGPAGKDGAGAGQGGPPGAPGAGERGEGQGA 477
 QY 169 GSPFGQLGPPAGPGEAGKGEQGVPGDLAGAPSGPAGE-----PGP----- 212
 DB 478 GSPFGQLGPPAGPGEAGKGEQGVPGDLAGAPSGPAGE-----PGP----- 537
 QY 213 -----TGLPGPGEGRGSGRGGPGADGAGVAGKPGAGRGSGPPGARGVQGGPPGAPGR 537
 DB 538 GAGAPGNDGAKGADGAPGAGSGQAGLQMPGEGRGAALPGPKDGRDAGPGADGSP 597
 QY 259 GE-----AGRPGAGLPGAK-----GLTSGSPGSPGDKTGGPPGAGQDGRP 300
 DB 598 GKDGVRGLTGTIPGPPAGAGDGGESGSPGAGTGAAGAPGDRGEPGPPGAGFAGPP 657
 QY 301 GPPGPPGARGQ-----AGVMGFPKGKGAAGFGKAGRGVPGPPCA----- 341
 DB 658 GADGQPGKAGBPGDAGAKGDAGPPGAPGAPGPPGPIGNVGAAPKAGARGSGAPGATGFP 717
 QY 342 -----VGPAGKDGAGAGGPPGAPG-----AGRGEGQAGSGFGQLGPPGAPGCEAGKP 393
 DB 718 GAAGRVGPPGSGNAGPPGPPGAPGKGGKGRGETGTGAPRGVGGPPGPPGAGEKGGSP 777
 QY 394 GEQGVPGDLGAPSGSPAGEPGTGLPCCGCGGSGRGGPGADGAGVAGKPGAGRGSGP 453
 DB 778 GADGAPAGPTGPGQIAGRGVGLPQORGERGPGPLPGSPGPKGSGSAGRGSP 837
 QY 454 GPAGP---KGSFGEAGRGAGLPGAKGLTSGSGSPGDKTGGPPGAGQDGRPPGPPG 510
 DB 838 GPMGPPGLAGPPGSGREGAPCAEGSPCRDGSPPAKGDRGTGTGAPGPPGAPGAPGAPV 897
 QY 511 GARGQAGVMGFPKGKAGRGVPGPGCAVGPAGKDG-----AGAG 558
 DB 898 GPAGKSGDRGTGTGAPGAPGPPGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 957
 QY 559 GPPGP-----AGPAGERGEGQAGSP---GFGQLGPPGAPGPPGAGKGEQGV 603
 DB 958 GPPGPPGSGGSGSAGPAGPPGSPAGAPKDGKGLNGLPGLPGLPGRGRTGDAGPV 1017
 QY 604 GDLGAPGSGPAG 616
 DB 1018 GPPGPPGPPGPPG 1030

RESULT 6
 AAY84540
 ID AAY84540 standard; protein; 1107 AA.
 XX
 AC AAY84540;
 XX
 DT 25-JUL-2000 (first entry)

XX Amino acid sequence of a chimeric collagen 1 (alpha1)/decorin protein.
 DE Extracellular matrix protein; self aggregation; hydroxylated proline;
 XX trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibrinogen; fibronectin; post translational hydroxylation;
 KW decorin; chimera.
 XX
 OS Homo sapiens.
 OS Unidentified.
 OS Chimeric.
 FH Key Location/Qualifiers
 FT Misc-difference 858 /note= "Gly encoded by GCT"
 XX
 FN EP992586-A2.
 XX
 PD 12-APR-2000.
 XX
 PF 07-OCT-1999; 99EP-00119184.
 XX
 PR 09-OCT-1998; 98US-00169768.
 XX
 PA (USSU) US SURGICAL CORP.
 XX
 XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
 PI WPI; 2000-259138/23.
 XX
 DR N-PSDB; AAA12500.
 XX
 XX Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX
 PS Claim 24; Fig 18; 260pp; English.
 XX
 CC The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a chimeric collagen 1 (alpha1)/decorin
 CC protein, which may be produced using the method of the invention
 XX
 SQ Sequence 1107 AA;

Query Match 64.1%; Score 2237.5; DB 3; Length 1107;
 Best Local Similarity 60.2%; Pred. No. 3.3e-130;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;
 QY 1 GPP-----GEPGPTGLPCCGCGGSGRGGPGADGAGVAGKPGAGRGSGPGA 48
 DB 297 GPPGAGBEGKRGARGEPGPTGLPCCGCGGSGRGGPGADGAGVAGKPGAGRGSGPGA 356
 QY 49 GPKSGPGEAGRGAGLPGAKGLTSGSGSPGDKTGGPGAGODGRPPGPPGARGQA 108
 DB 357 GPKSGPGEAGRGAGLPGAKGLTSGSGSPGDKTGGPGAGODGRPPGPPGARGQA 416
 QY 109 GVMGFPKGKGAAGFGKAGRGVPGPGCAVGPAGKDGAGAGQGGPPGAPGAGERGEGQGA 168

Db 938 GADPAGAGTGGGQAGQVGVVGLPQGRGFFGLPGSPGPKQSGASGERGPP 997
Qy 454 GPAGP---KGSPGAGRPGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPP 510
Db 998 GPMGPPGLAGPPGSGREGAPGAGSGPGRDGSFGAKGDRGTGPPGAGPAGPAGPV 1057
Qy 511 GARGQAGVMGPPGPKGAAGBPGKAGRGVPPGAVGPAKQGE-----AGAQ 558
Db 1058 GPAGKSGDRGTGPPAGPAGVGVGAGPAGPAGQPRGDKGTGECQDRIKGRHSGSLQ 1117
Qy 559 GPPGP-----AGPAGRGQGPAGSP---GFOGLPAGPAGPAGPAGPAGV 603
Db 1118 GPPGPPGSGPQSGASGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1177
Qy 604 GDLGAPGSPGAG 616
Db 1178 GPPGPPGPPGPPG 1190

RESULT 9
ADD45059
ID ADD45059 standard; protein; 1464 AA.
XX AC ADD45059;
XX AC
DT 29-JAN-2004 (first entry)
XX DE Human Protein P02452, SEQ ID NO 10491.
XX KW Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX OS Homo sapiens.
OS
PN WO2003016475-A2.
PN
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GEO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WPI: 2003-268312/26.
DR GENBANK; P02452.
DR
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
PS Claim 1; Page; 1017pp; English.
PS
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a

CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1464 AA;

Query Match 64.1%; Score 2237.5; DB 7; Length 1464;
Best Local Similarity 60.2%; Pred. No. 4.1e-130;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;
QY 1 GPP-----GPPGPTGLPSPGSGRGGPAGADGVAGPKGPPGAGRGGPQA 48
Db 458 GPPGAGBEGKRGARGEPGPTGLPSPGSGRGGPAGADGVAGPKGPPGAGRGGPQA 517
QY 49 GPKGSPGAGRPGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGAGQA 108
Db 518 GPKGSPGAGRPGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGAGQA 577
QY 109 GVMGPPGKGAAGBPGKAGRGVPPGAVGPAKQGEAGAQDGRPPGAGRGGPQA 168
Db 578 GVMGPPGKGAAGBPGKAGRGVPPGAVGPAKQGEAGAQDGRPPGAGRGGPQA 637
QY 169 GSPGFQGLPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 212
Db 638 GSPGFQGLPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 697
QY 213 -----TGLPSPGSGRGGPAGADGVAGPKGPPGAGRGGPAGRGGPQA 258
Db 698 GAGAPGNDGAKGADAGAPGAPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 757
QY 259 GE-----AGRPGAGLPGAK-----GLTSPGSPGPDGKTGPPGAGQDGRP 300
Db 758 GKGVRLGTGPIGPPGAGAPGDKGSGSPGAPGTGARGAPGDRGEPGPPGAGFAGPP 817
QY 301 GPPGPPGARGQ-----AGVMGPPGKGAAGBPGKAGRGVPPGPPGA--- 341
Db 818 GADQPGAKGEPGDAGAKGADGAPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 877
QY 342 -----VGPAGKDEAGAGQPPGAGP-----AGERGEGPAGSPGFGGLPAGPAGPAG 393
Db 878 GAAGRVPFPSPGNAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 937
QY 394 GEOGVPGDLGAPGSPGAPGPTGLPSPGSGRGGPAGADGVAGPKGPPGAGRGGP 453
Db 938 GAGGAPAGPPTPQGLAGQGVVGLPQGRGEGFPGLPGSPGEPKQSGASGERGPP 997
QY 454 GPAGP---KGSPGAGRPGAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPP 510
Db 998 GPMGPPGLAGPPGSGREGAPGAGSGPGRDGSFGAKGDRGTGPPGAGPAGPAGPV 1057
QY 511 GARGQAGVMGPPGPKGAAGBPGKAGRGVPPGAVGPAKQGE-----AGAQ 558
Db 1058 GPAGKSGDRGTGPPAGPAGVGVGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 1117
QY 559 GPPGP-----AGPAGRGQGPAGSP---GFOGLPAGPAGPAGPAGPAGV 603
Db 1118 GPPGPPGSGPQSGASGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1177
QY 604 GDLGAPGSPGAG 616
Db 1178 GPPGPPGPPGPPG 1190

Query Match	64.1%	Score 2337.5	DB 7	Length 1464
Best Local Similarity	60.2%	Pred. No. 4.1e-130		
Matches 441	Conservative 23	Mismatches 152	Indels 117	Gaps 12

SQ Sequence 1464 AA;

Query Match 64.1%; Score 2237.5; DB 8; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 4.1e-130;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGPPGCGSGRGGPGGADGVAGPKGAGRGSGCPGA 48
 DB 458 GPPGAGEGKRGARGEPGPTGLPGPPGCGSGRGGPGGADGVAGPKGAGRGSGCPGA 517

QY 49 GPKSGPGEAGRPGAGLPGAKGLTSGSGSPGPDGKTGPPGAGQDGRGPPGPPGARGQA 108
 DB 518 GPKSGPGEAGRPGAGLPGAKGLTSGSGSPGPDGKTGPPGAGQDGRGPPGPPGARGQA 577

QY 109 GVMGPPGKGAAGPGEAGRGVPGPCAVGAGKDGAGAGQGGPPGAGPAGERGEGQPA 168
 DB 578 GVMGPPGKGAAGPGEAGRGVPGPCAVGAGKDGAGAGQGGPPGAGPAGERGEGQPA 637

QY 169 GSPGFGGLPGPAGPGEAGRGVPGPCAVGAGKDGAGAGQGGPPGAGPAGERGEGQPA 212
 DB 638 GSPGFGGLPGPAGPGEAGRGVPGPCAVGAGKDGAGAGQGGPPGAGPAGERGEGQPA 697

QY 213 -----TGLPGPPGCGSGRGGPGGADGVAGPKGAGRGSGCPGAPGKSP 258
 DB 698 GAGAPGNDGAKGADGAGPAGPSOGAPGQGMPEGERGAAGLPGKGDGADGPKGADGSP 757

QY 259 GE-----AGRPGAGLPGAK-----GLTSGSPSGPDGKTGPPGAGQDGRGPP 300
 DB 758 GKDGVRGLTGTPIGPPGAGAPGDKGESGSPGAGTGAAGAPGDRGEPGPPGAGFAGPP 817

QY 301 GPPGPPGARGQ-----AGVMGFPKGAAGRGKAGRGVPGPCA----- 341
 DB 818 GADGQPKAGKGPBGDAGAKGDAGPPGAPGPPGPGIENVGAPGAKGARGSGAPGATGFP 877

QY 342 -----VGPAGKDGAGAGQGGPPGAGP---AGERGEQGPAGSPGQGLPGPAGPGEAGKP 393
 DB 878 GAAGRVGPPGSGNAGPPGPPGPKAGKGGKPRGETGAPRGVGEVPPGPPGAGEKSP 937

QY 394 GEQGVPGDLCAPGSGPAGPAGTGLPGPPGCGSGRGGPPGADGVAGPKGAGRGSGP 453
 DB 938 GADGAPAGAPGTPGPGIAGQGVVGLPGQGRGERGFPGLPGSPSGPKQKQPSGASGERGPP 997

QY 454 GPAGP---KGSPPGAGRPGGAGLPGAKGLTSGSGSPGPDGKTGPPGAGQDGRGPP 510
 DB 998 GPMGPPGLAGPPGSGREGAGPCAGSGPRDGSPPGAKGDRGETGAGPPGAPGAPGAPGV 1057

QY 511 GARGQAGVMGPPGPKGAAGEPKAGRGVPGPCAVGAGKDG-----AGAQ 558
 DB 1058 GPAGKSGDRGTGTPAGPAGPVGPPGARGPAGPQGGPRDKGTGEGDRGIKGRGFSGLQ 1117

QY 559 GPPGP-----AGPAGERGEGQGPAGSP---GFQGLPGPAGPGEAGKPGEGQGV 603
 DB 1118 GPPGPPGSGGSGSAGPAGPRGPPGSGAGAPGKDGGLNGLPGIPIGPPGRTGDAGPV 1177

QY 604 GDLGAPGSPGAG 616
 DB 1178 GPPGPPGPPGPPG 1190

RESULT 13
 ADRI6425
 ID ADRI6425 standard; protein; 1464 AA.
 XX AC ADRI6425;
 XX DT 21-OCT-2004 (first entry)
 XX DE Human collagen I alpha1 (I) chain protein.
 XX KW T-lymphocyte; AAA; abdominal aortic aneurysm; vaccine; therapy; human;
 XX KW collagen I.
 XX OS Homo sapiens.

XX EN US2004151732-A1.
 XX PD 05-AUG-2004.
 XX PF 04-FEB-2003; 2003US-00358024.
 XX PR 04-FEB-2003; 2003US-00358024.
 XX PA (JICH// JICHA D L.
 XX PA (PELU//) PELUSE S.
 XX PI Jicha DL, Peluse S;
 XX DR WPI; 2004-570707/55.
 XX PT Isolated or purified lymphocytes derived from blood in abdominal aortic
 PT aneurysm patients useful for early diagnosis of aneurysms are reactive
 PT with collagen I, collagen III and/or their fragments.
 XX FS Disclosure; SEQ ID NO 1; 22pp; English.
 XX CC The present provides a T-lymphocyte derived from blood in abdominal
 CC aortic aneurysm (AAA) patients which are reactive with collagen I,
 CC collagen III and their fragments. The invention is useful in diagnosis,
 CC prevention and treatment of abdominal aortic aneurysm or rupture in a
 CC mammal. The invention is also useful in the vaccine preparation. The
 CC present sequence is human collagen I alpha1 (I) chain protein.
 XX SQ Sequence 1464 AA;

Query Match 64.1%; Score 2237.5; DB 8; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 4.1e-130;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGPPGCGSGRGGPGGADGVAGPKGAGRGSGCPGA 48
 DB 458 GPPGAGEGKRGARGEPGPTGLPGPPGCGSGRGGPGGADGVAGPKGAGRGSGCPGA 517

QY 49 GPKSGPGEAGRPGAGLPGAKGLTSGSGSPGPDGKTGPPGAGQDGRGPPGPPGARGQA 108
 DB 518 GPKSGPGEAGRPGAGLPGAKGLTSGSGSPGPDGKTGPPGAGQDGRGPPGPPGARGQA 577

QY 109 GVMGPPGKGAAGPGEAGRGVPGPCAVGAGKDGAGAGQGGPPGAGPAGERGEGQPA 168
 DB 578 GVMGPPGKGAAGPGEAGRGVPGPCAVGAGKDGAGAGQGGPPGAGPAGERGEGQPA 637

QY 169 GSPGFGGLPGPAGPGEAGRGVPGPCAVGAGKDGAGAGQGGPPGAGPAGERGEGQPA 212
 DB 638 GSPGFGGLPGPAGPGEAGRGVPGPCAVGAGKDGAGAGQGGPPGAGPAGERGEGQPA 697

QY 213 -----TGLPGPPGCGSGRGGPGGADGVAGPKGAGRGSGCPGAPGKSP 258
 DB 698 GAGAPGNDGAKGADGAGPAGPSOGAPGQGMPEGERGAAGLPGKGDGADGPKGADGSP 757

QY 259 GE-----AGRPGAGLPGAK-----GLTSGSPSGPDGKTGPPGAGQDGRGPP 300
 DB 758 GKDGVRGLTGTPIGPPGAGAPGDKGESGSPGAGTGAAGAPGDRGEPGPPGAGFAGPP 817

QY 301 GPPGPPGARGQ-----AGVMGFPKGAAGRGKAGRGVPGPCA----- 341
 DB 818 GADGQPKAGKGPBGDAGAKGDAGPPGAPGPPGPGIENVGAPGAKGARGSGAPGATGFP 877

QY 342 -----VGPAGKDGAGAGQGGPPGAGP---AGERGEQGPAGSPGQGLPGPAGPGEAGKP 393
 DB 878 GAAGRVGPPGSGNAGPPGPPGPKAGKGGKPRGETGAPRGVGEVPPGPPGAGEKSP 937

QY 394 GEQGVPGDLCAPGSGPAGPAGTGLPGPPGCGSGRGGPPGADGVAGPKGAGRGSGP 453
 DB 938 GADGAPAGAPGTPGPGIAGQGVVGLPGQGRGERGFPGLPGSPSGPKQKQPSGASGERGPP 997

QY 454 GPAGP---KGSPPGAGRPGGAGLPGAKGLTSGSGSPGPDGKTGPPGAGQDGRGPP 510
 DB 998 GPMGPPGLAGPPGSGREGAGPCAGSGPRDGSPPGAKGDRGETGAGPPGAPGAPGAPGV 1057

QY 511 GARGQAGVMGPPGPKGAAGEPKAGRGVPGPCAVGAGKDG-----AGAQ 558
 DB 1058 GPAGKSGDRGTGTPAGPAGPVGPPGARGPAGPQGGPRDKGTGEGDRGIKGRGFSGLQ 1117

QY 559 GPPGP-----AGPAGERGEGQGPAGSP---GFQGLPGPAGPGEAGKPGEGQGV 603
 DB 1118 GPPGPPGSGGSGSAGPAGPRGPPGSGAGAPGKDGGLNGLPGIPIGPPGRTGDAGPV 1177

QY 604 GDLGAPGSPGAG 616
 DB 1178 GPPGPPGPPGPPG 1190

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:46:04 ; Search time 26.1894 Seconds
(without alignments)
1758.668 Million cell updates/sec

Title: US-10-658-989A-2
Perfect score: 3488
Sequence: 1 GPPGEPGTGLPGRPGRGG.....GEOGVPGDLGAPGSPAGG 617

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pap:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pap:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pap:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pap:*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pap:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2237.5	64.1	1464	4	US-09-331-347C-21
2	2228.5	63.9	1341	3	US-08-963-825-18
3	2228.5	63.9	1341	3	US-09-500-811-18
4	2228.5	63.9	1341	3	US-09-570-573-18
5	2228.5	63.9	1341	3	US-09-548-608-18
6	2224.5	63.8	1461	4	US-09-585-887-9
7	2224.5	63.8	1461	4	US-09-289-578-9
8	2208.5	63.3	1057	3	US-08-931-820-1
9	2191	62.8	822	3	US-09-213-849-49
10	1955	56.0	1017	4	US-08-468-996-10
11	1943	55.7	1060	3	US-08-931-820-3
12	1943	55.7	1418	3	US-08-963-825-20
13	1943	55.7	1418	3	US-09-101-999-1
14	1943	55.7	1418	3	US-09-500-811-20
15	1943	55.7	1418	3	US-09-570-573-20
16	1943	55.7	1418	3	US-08-316-650-12
17	1936	55.5	1442	2	US-08-316-650-12
18	1936	55.5	1442	5	PCT-US95-02251-12
19	1882.5	54.0	595	3	US-09-219-849-48
20	1882.5	54.0	595	3	US-09-219-849-50
21	1882.5	54.0	1064	1	US-08-642-255-62
22	1853	53.1	1057	3	US-08-931-820-4
23	1837	52.7	720	3	US-09-219-849-4
24	1837	52.7	777	1	US-08-642-255-53
25	1833.5	52.6	1078	3	US-08-963-825-21
26	1833.5	52.6	1078	3	US-09-500-811-21
27	1833.5	52.6	1078	3	US-09-570-573-21

28	1833.5	52.6	1078	3	US-09-548-608-21	Sequence 21, Appl
29	1764	50.6	1024	3	US-08-931-820-2	Sequence 2, Appl
30	1764	50.6	1366	3	US-08-963-825-19	Sequence 19, Appl
31	1764	50.6	1366	3	US-09-500-811-19	Sequence 19, Appl
32	1764	50.6	1366	3	US-09-570-573-19	Sequence 19, Appl
33	1764	50.6	1366	3	US-09-548-608-19	Sequence 19, Appl
34	1762	50.5	1366	4	US-09-585-887-10	Sequence 10, Appl
35	1762	50.5	1366	4	US-09-289-578-10	Sequence 10, Appl
36	1762	50.5	1366	4	US-09-949-016-5882	Sequence 5882, Ap
37	1699.5	48.7	1065	1	US-08-642-255-72	Sequence 72, Appl
38	1674	48.0	960	3	US-09-219-849-5	Sequence 5, Appl
39	1645	47.2	492	4	US-08-468-996-12	Sequence 12, Appl
40	1644	47.1	1806	4	US-09-919-497-56	Sequence 56, Appl
41	1619.5	46.4	829	1	US-08-642-255-132	Sequence 132, App
42	1619.5	46.4	829	1	US-08-397-633A-53	Sequence 53, Appl
43	1619.5	46.4	837	1	US-08-175-155-68	Sequence 68, Appl
44	1619.5	46.4	837	1	US-08-477-509B-103	Sequence 103, App
45	1619.5	46.4	837	1	US-08-642-255-101	Sequence 101, App

ALIGNMENTS

RESULT 1
US-09-331-347C-21
; Sequence 21, Application US/09331347C
; Patent No. 6617431
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Met
; TITLE OF INVENTION: obtaining Such and Their Uses
; FILE REFERENCE: 1149-3
; CURRENT APPLICATION NUMBER: US/09/331.347C
; CURRENT FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-331-347C-21

Query Match	64.1%	Score	2237.5	DB	4	Length	1464
Best Local Similarity	60.2%	Pred. No.	1.3e-133				
Matches	441	Conservative	23	Mismatches	152	Indels	117
Gaps	12						
Qy	1	GPP-----	-----GEGPTGLPGRGGRGPGGSRGPGGADGVAGPKGAGERGSPGPA	48			
Db	458	GPPGAGEEKGARGEPGPTGLPGRGGRGPGGSRGPGGADGVAGPKGAGERGSPGPA	517				
Qy	49	GPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGDKTGPFGAGQDGRFPFGPGARQQA	108				
Db	518	GPKGSPGEAGRPGEAGLPGAKGLTGSPGSPGDKTGPFGAGQDGRFPFGPGARQQA	577				
Qy	109	GVMGFPKGAAGPFGKAGRGVPGPCAVGPAKDGAGAGQPPGAPGAGERGQCPA	168				
Db	578	GVMGFPKGAAGPFGKAGRGVPGPCAVGPAKDGAGAGQPPGAPGAGERGQCPA	637				
Qy	169	GSPGFQGLPGPAGPGEAGKGEQGVCDLCAAGPSGPAE-----	212				
Db	638	GSPGFQGLPGPAGPGEAGKGEQGVCDLCAAGPSGPAE-----	697				
Qy	213	-----TGLPGRGGRGPGSRGPFPGADGVAGPKGAGERGSPGPKGSP	258				
Db	698	GANGAPGNDGAKGADGAFAGFQAGLQGMFGERGAGLPGKGDGADGPKGADSP	757				
Qy	259	GE-----AGRPGEAGLPGAK-----GLTSPGSPGDKTGPFGPAGQDGRP	300				
Db	758	GKDGVRGLTGTPGPPGAPAGDKGESGSPGAPGTGARGAPGDRGEPFGPAGFAGPP	817				
Qy	301	GPPGPGARGO-----AGVMGFPKGAAGPFGKAGRGVPGPCA-----	341				

NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/09/500.811
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/187,319
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)
US-09-500-811-18

Query Match 63.9%; Score 2228.5; DB 3; Length 1341;
Best Local Similarity 60.0%; Pred. No. 4.6e-133;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

QY 1 GPP-----GEPCTGLPGERGPGSRGPPGADGVAGPKGAGERGSPGA 48
|||
Db 334 GPPGAGEGKRGARGEFGTGLPGERGPGSRGPPGADGVAGPKGAGERGSPGA 393
QY 49 GPKSGPAGRPGEGAGLFGAKGLTSGSPGPDGKTGPPGAGQDGRGPPGARGQA 108
|||
Db 394 GPKSGPAGRPGEGAGLFGAKGLTSGSPGPDGKTGPPGAGQDGRGPPGARGQA 453
QY 109 GVMGPPGKGAAGEPKAGRGVPCGAVGPGAGKDGAGAGPPGAPGAGERGEOGPA 168
|||
Db 454 GVMGPPGKGAAGEPKAGRGVPCGAVGPGAGKDGAGAGPPGAPGAGERGEOGPA 513
QY 169 GSPFGQLGPPAGPGEAGKPGEGQVPGDLGAPGSPGAGE-----PGP---- 212
|||
Db 514 GSPFGQLGPPAGPGEAGKPGEGQVPGDLGAPGSPGAGERGPPGGERGVQPPGAPGR 573
QY 213 -----TGLPGPGERGPGSRGPPGADGVAGPKGAGERGSPGAPGKSP 258
|||
Db 574 GANPAGNDGAKGDAGAPGAGSQAGLQGMFGERGAAGLFGPKGDRGDAGPKGADGSP 633
QY 259 GE-----AGRGEAGLPGAK-----GLTSGSPGPDGKTGPPGAGQDGRP 300
|||
Db 634 GKDGVRLGTGTPGPPGAPAGDKGESGSPGAGTGAAGPDRGEPGPPGAPGAPGP 693
QY 301 GPPGPPGARGQ-----AGVMGPPGKGAAGEPKAGRGVPCGPA----- 341
|||
Db 694 GADGQPGAKGEGFDAGAKGDAGPPGAPGAPGPPGIGNVGAFGAKGARGSPGATGTF 753
QY 342 -----VGPAGKDGAGAGQPPGAPG-----AGRGEGQPGAGSPFGQLGPPGAPGGEAGKP 393

Db 754 GAAGRVGPPGSPGNAGPPGPPGAGKGGKGRGETGPAAGPGEVPPGPPGAGEKSP 813
QY 394 GEQVPGDLGAPGSPGAPGPPGTLGPPGGERGPGSRGPPGADGVAGPKGAGERGSP 453
Db 814 GADGAPAGPTGPPGQIAGQGVVGLPGQGERGPPGGLPGSPGPKQGPSGASGERGPP 873
QY 454 GPAGP---KSGPGEAGRPGEGAGLFGAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPP 510
Db 874 GPMGPPGLAGPPGESGREGAPGAEPSGRDGSPPAKGDRGETGAPGPPGAXGAGAPGV 933
QY 511 GARGQAGVMGPPGPKGAAGEPKAGRGVPCGAVGPGAGKDG-----AGAQ 558
Db 934 GPACKSGDRGETGAPGAPGVPVGPAGARGPAGPQGRGDKGTGQDGRGIKGRHGFSLQ 993
QY 559 GPPGP-----AGPAGERGEOGPPAGSP--GFQGLGPPGAPGAGERGKGEQGV 603
Db 994 GPPGPPGSPGEGQSPGASGAPGPRGPPGASAGAPCKDGLGGLPGIPGPPGRRTGDAGPV 1053
QY 604 GDLGARGSPGAP 616
Db 1054 GPPGPPGPPGPPG 1066

RESULT 4

US-09-570-573-18
Sequence 18, Application US/09570573
Patent No. 6342361

GENERAL INFORMATION:
APPLICANT: Ovist, Per

TITLE OF INVENTION: A Method for Assaying Collagen Fragments

TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the

TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

TITLE OF INVENTION: Disorders Associated with the Metabolism of

NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC

STREET: 805 Third Avenue
CITY: New York

STATE: New York
COUNTRY: USA

ZIP: 10022
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA: US/09/570.573
APPLICATION NUMBER: 08/187,319

FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA: 08/187,319
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C

REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700

TELEFAX: 212-753-6237
TELEX: 236687

INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:

LENGTH: 1341 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

ORIGINAL SOURCE:
ORGANISM: Homo sapiens

IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)


```
QY 394 GEQVPGDLGAPGSPGAGEGPTGLPGPCGCGSGRGPFGADGVAGPKGAPAGERGSP 453
Db 814 GADGAPAGAGPTGPGQIAGQGVVGLPGQRGEPFGLPGSGGPGKQSGASGERGPP 873
QY 454 GPAGP---KGSFGBAGRGEAGLPCAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPP 510
Db 874 GPMGPPGLAGFPFSGREGAPGAEBSFGSDGSPGAKGDRGETGAGPPGAXGAXGAPGV 933
QY 511 GARGQAGVMGFPFGPKGAAGEFGKAGRGVPPGPAAGVAGPKADGE-----AGAQ 558
Db 934 GPAGKSGDRGETGAPGAPGVPAGAPGAPGQPRGDKGTGEGQDRIKGRHGFSLQ 993
QY 559 GPPGP-----AGPAGERGEOGAGSP---GFGGLPGAGPPGEAGKGEQGV 603
Db 994 GPPGPPGSGQSGASGAPGPRGPPGASAGPKDGLNGLPGIPGPPRGRGTGDAGPV 1053
QY 604 GDLGAPGSPGAG 616
Db 1054 GPPGPPGPPGPPG 1066

RESULT 6
US-09-585-887-9
; Sequence 9, Application US/09585887
; Patent No. 6413742
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/585,887
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/289,578
; PRIOR FILING DATE: 1999-04-09
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-585-887-9

Query Match 63.8%; Score 2224.5; DB 4; Length 1461;
Best Local Similarity 59.9%; Pred. No. 8.8e-133;
Matches 439; Conservative 23; Mismatches 154; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGPCGSGRGPFGADGVAGPKGAPAGERGSPCPA 48
Db 455 GPPGAGEGKRGAGEGPTGLPGPCGSGRGPFGADGVAGPKGAPAGERGSPCPA 514
QY 49 GPKSGPGEAGRGEAGLPCAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPGARQQA 108
Db 515 GPKSGPGEAGRGEAGLPCAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPGARQQA 574
QY 169 GSPGQGLPGAGPPGAGEGKAGEQGVPCDILGAPGSPGAGE-----GPG----- 212
Db 635 GSPGQGLPGAGPPGAGEGKAGEQGVPCDILGAPGSPGAGEGKAGEQGVPCDILGAPG 694
QY 213 -----TGLPGPCGCGSGRGPFGADGVAGPKGAPAGERGSPGAPGKPS 258
Db 695 GAGAPGNDGAKGDAGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 754
```

```
QY 259 GE-----AGRPGEAGLPGAK-----GLTSGPSGPDGKTGPPGAGQDGRP 300
Db 755 GKQGVRLGTPIPGPPGAPAGDKGESGSPGAPGPTGARGAGDGRGEPGPPGAPGAPPP 814
QY 301 GPPGPPGARGQ-----AGVMGFPKGAAGEGPKGAGERGVPGPPGA----- 341
Db 815 GADQPCAKGEPGDAGAKGDAGPPGAPGAPGPPGICNVGAPGAKGARGAGSAGPPGATGFP 874
QY 342 -----VCPAGKDEAGAGQPPGAPG---AGERGEOGAPGSPGQGLPGAGPPGAGKP 393
Db 875 GAAGRVPFPFSGNAGPPGPPGAGKEGKGRGETGAPRGPCGEVGPVPPGPPGAGERGSP 934
QY 394 GEQVPCDLGAPGSPGAPGPPGPTGLPGPCGCGSGRGPFGADGVAGPKGAPAGERGSP 453
Db 935 GADGAPAGAGTTPGQIAGQGVVGLPGQRGEPFGLPGSGGPGKQSGASGERGPP 994
QY 454 GPAGP---KGSFGBAGRGEAGLPCAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPP 510
Db 995 GPMGPPGLAGFPFSGREGAPGAEBSFGSDGSPGAKGDRGETGAGPPGAXGAXGAPGV 1054
QY 511 GARGQAGVMGFPFGPKGAAGEFGKAGRGVPPGPAAGVAGPKADGE-----AGAQ 558
Db 1055 GPAGKSGDRGETGAPGAPGVPVAGAPGAPGQPRGDKGTGEGQDRIKGRHGFSLQ 1114
QY 559 GPPGP-----AGPAGERGEOGAGSP---GFGGLPGAGPPGEAGKGEQGV 603
Db 1115 GPPGPPGSGQSGASGAPGPRGPPGASAGPKDGLNGLPGIPGPPRGRGTGDAGPV 1174
QY 604 GDLGAPGSPGAG 616
Db 1175 GPPGPPGPPGPPG 1187

RESULT 7
US-09-289-578-9
; Sequence 9, Application US/09289578
; Patent No. 6428978
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/289,578
; PRIOR FILING DATE: 1999-04-10
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-289-578-9

Query Match 63.8%; Score 2224.5; DB 4; Length 1461;
Best Local Similarity 59.9%; Pred. No. 8.8e-133;
Matches 439; Conservative 23; Mismatches 154; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGPCGSGRGPFGADGVAGPKGAPAGERGSPCPA 48
Db 455 GPPGAGEGKRGAGEGPTGLPGPCGSGRGPFGADGVAGPKGAPAGERGSPCPA 514
QY 49 GPKSGPGEAGRGEAGLPCAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPGARQQA 108
Db 515 GPKSGPGEAGRGEAGLPCAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPGARQQA 574
QY 109 GVMGFPFGKGAAGEFGKAGRGVPPGPAAGVAGPKGAPAGERGSPGAPGKPS 168
Db 695 GAGAPGNDGAKGDAGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 754
```

```

Db 575 GVMGFPKGAAGEPKAGRGVPPGAVGAGKDEAGAGQGGPPGAPGAGERGEQGP 634
QY 169 GSPFGQGLPGAPGPEAGKGEQGVPGDLGAPGSPGAGE-----PGP----- 212
Db 635 GSPFGQGLPGAPGPEAGKGEQGVPGDLGAPGSPGAGERGPPGGERGVGPPGAPGR 694
QY 213 -----TGLPGPPEGRGGPSRPFPGADGVAGPKGAGERSGSPGAPGKGS 258
Db 695 GANGAPGNDGAKDAGAPGAGSGAGLQGMFGERGAAGLPGPKGRDAGPKGADGSP 754
QY 259 GE-----AGRPGEAGLPGAK-----GLTGSFGSPGDKTGTGPPGAGQGRP 300
Db 755 GKDGVRLGTGPIGPPGAPGDKGESGSPGAPGPTGARGAPGDRGFPFGPPGAFAGPP 814
QY 301 GPPGPPGARGQ-----AGVMGFPKGAAGEPKAGRGVPPGPGA----- 341
Db 815 GADGQFGAKPGDAGAKDAGPPGAPGAPGPPGPIGNVGAAPGAKGAGSGPPGATGFP 874
QY 342 -----VGPAGKDGAGAGPPGAPG-----AGERGEQGPAGSPGQGLPGAPGPPGAGKP 393
Db 875 GNAAGRVGPPSGNAGPPGPPGAPGKGGKGRGETGAPRGVGVGPPGPPGAPGKGS 934
QY 394 GEQGVPGDLGAPGSPGAPGPTGLPGPFGERGPGSRGFPAGADGVAGPKGAPGERSG 453
Db 935 GADGPAGAPGTPGQIAGQGVVGLPGQRGGRFPGLPGPSEPKGQSGASGERGPP 994
QY 454 GPAGP---KSGSPGAGRPEAGLPGAKGLTGSFGSPGDPDKTGTGPPGAGQDGRPPGPP 510
Db 995 GPMGPPGLAGPPGSGREGAFGAEFGSPGRDGPAGKGRDGETGAPGAPGAPVAPGV 1054
QY 511 GARGQAGVMGPPGPKGAAGEPKAGRGVPPGAVGAPGAKDGE-----AGAQ 558
Db 1055 GPAGKSGDRGRTGAPGAPGVGPGARGAPGQPGRGDKGETGEQDRIKGRHGFSLQ 1114
QY 559 GPRGP-----AGPAGERGEQGPAGSP---GFQGLPGPAGPGEAGKPGESQGV 603
Db 1115 GPPGPPSGEQGPGSGASGAPGRGPPSGAGAPGKDLNGLPGPIGPPGRGRTGDAGPV 1174
QY 604 GDLGAPGSPGAP 616
Db 1175 GPPGPPGPPGPPG 1187

```

RESULT 8

```

US-08-931-820-1
; Sequence 1, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:

```

```

; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type I
; US-08-931-820-1

```

```

Query Match 63.3%; Score 2208.5; DB 3; Length 1057;
Best Local Similarity 60.3%; Pred. No. 6.9e-132;
Matches 435; Conservative 23; Mismatches 158; Indels 105; Gaps 11;

```

```

QY 1 GPPGPPGTGLPGPFGGSGRFPAGDGVAGPKGAPGERSGPPAGPKSGSPGACRP 60
Db 309 GPNGEAGSAGPPGPPGLRGSGRFPAGDGVAGPKGAPGERSGPPAGPKSGSPGACRP 368
QY 61 GEAGLPGAKGLTGSFGSPGDPGKTGPPGAGQDGRPPGPPGARGQAGVMGFPKGA 120
Db 369 GEAGLPGAKGLTGSFGSPGDPGKTGPPGAGQDGRPPGPPGARGQAGVMGFPKGA 428
QY 121 GEPGKAGRGVPPGAVGAPGAGKDEAGAGQGPAGPAGRGEGGAGSGPQGLPGPA 180
Db 429 GEPGKAGRGVPPGAVGAPGAGKDEAGAGQGPAGPAGRGEGGAGSGPQGLPGPA 488
QY 181 GPPGAGTGGGQGVPGDLGAPGSPGAGE-----PGP----- 212
Db 489 GPPGAGTGGGQGVPGDLGAPGSPGAGERGFPGERGVQGGPPGAPRGANGAPGNDGAK 548
QY 213 ---TGLPGPPEGRGGPSRPFPGADGVAGPKGAPGERSGPPGAPGKGSFGE-----A 261
Db 549 GDAGAPGAPGSGQAGLPGQGMFGERGAAGLPGPKGRDAGPKGADGSPGKDVRLGTPI 608
QY 262 GRPGEAGLPGAK-----GLTGSFGSPGDPDKTGTGPPGAGQDGRPPGPPGARGQ- 311
Db 609 GPPGAPGAPGDKGESGSPGAPGTGARGAPGDRGERGPPGAPGAGPPGADGPPGAKGEP 668
QY 312 -----AGVMGFPKGAAGEPKAGRGVPPGPGA-----VGPAGKD 348
Db 669 GDAGAKGDAGPPGAPGAPGPPGPIGNVGAAPGAKGARGSAGPPGATGFPGAAGRVPSPGS 728
QY 349 GEAGAGPPGPPAGP---AGERGEQGPAGSPGQGLPGPAGPGEAGKPGQGVPGDLGAP 405
Db 729 GNAGPPGPPGAPGKGGKGRGETGAPRGVGVGPPGPPGAPGKSGPGADGAPGAPGTP 788
QY 406 GPSGAPGEPGTGLPGPPEGRGGPSRPFPGADGVAGPKGAPGERSGPPGAP---KGSP 462
Db 789 GPQIAGQGVVGLPQQRGERGFPGLPGFSGEPGKQSGASGERGPPGPMGPPGLAGPP 848
QY 463 GEAGRPGEAGLPGAKGLTGSFGSPGDPDKTGTGPPGAGQDGRGPPGPPGARGQAGVMGPP 522
Db 849 GESREGAPGAEKSGPGRDGSFGAKGDRGETGAPGAPGAPGAPGAPGAPGAPGAPGAP 908
QY 523 GPKGAAGEPKAGRGVPPGAVGAPGAKDGE-----AGAQGPPG----- 563
Db 909 GPAGPAGVPVGAARGAPGAPGPPGQGRGDKGETGEQDRIKGRHGFSLQGGPPGSPGQ 968
QY 564 -----AGPAGERGEQGPAGSP---GFQGLPGPAGPGEAGKPGEQGVPGDLGAPGSPGA 615
Db 969 GPSGASGAPGPPGSPGASAGPKDGLNGLPGPIGPPPRGRTGDAGPVGPPGPPGPP 1028
QY 616 G 616
Db 1029 G 1029

```

RESULT 9

```

US-09-219-849-49
; Sequence 49, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWSTRAN, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOREK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.

```

APPLICANT: VAN DEN BOSCH, TANJA J.
TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
TITLE OF INVENTION: PREPARATION THEREOF
FILE REFERENCE: 2728-2
CURRENT APPLICATION NUMBER: US/09/219,849
CURRENT FILING DATE: 1998-12-23
NUMBER OF SEQ ID NOS: 50
SOFTWARE: Patent in ver. 2.1
SEQ ID NO 49
LENGTH: 822
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
OTHER INFORMATION: amino acid sequence
US-09-219-849-49

Query Match 62.8%; Score 2191; DB 3; Length 822;
Best Local Similarity 56.9%; Pred. No. 7e-131;
Matches 436; Conservative 22; Mismatches 150; Indels 150; Gaps 13;

Qy 1 GPPGPPGPTGLPPGGE-----RGCGSRGPPGADGVAGPKGPPAGERGSPGPA 48
Db 12 GPPGAGPQGGPGEGERGCGSGPMGPRGPPGPKNGDDGEAGKGRGERGPPGPQ 71
Qy 49 GPKGSPGAGRRP-----GEAGLPKAGLGTGSPGPPGDKTGGPPGAGQD 93
Db 72 GARGLPGLTAGLPGMKGHRGFSGLDCAKGDAGPAGPKGPPGSPGENGAPQMGPRGLPGR 131
Qy 94 GRPDPGPPGARGQGVGPPGPKAAGE-----PGKAGRGVPPGAVGPPAGKDEA 147
Db 132 GRPDPGPPGARGQGVGPPGPKAAGE-----PGKAGRGVPPGAVGPPAGKDEA 191
Qy 148 GAQGGPPGPPGAGRGVPPGPKAAGE-----AGSPGFGQLGPPGAPGPEAGKPCQE 192
Db 192 GEPGPPGPPGAGRGVPPGPKAAGE-----AGSPGFGQLGPPGAPGPEAGKPCQE 251
Qy 193 GVPGLDGLAP-----GPGTGLPFGPPGRRGP 225
Db 252 GNSGPPGAPGKNDGTAKGEGCATGVCQGPAGGEGKRGARGEPGSLGPPGRRGP 311
Qy 226 GSRGPPGADGVAGPKGPPAGERGSPGAPGKSPGPPGAGRPBAGLPGAKGLTSGSPGPPD 285
Db 312 GSRGPPGADGVAGPKGPPAGERGSPGAPGKSPGPPGAGRPBAGLPGAKGLTSGSPGPPD 371
Qy 286 GRTGPPGAGQDGRPPGPPGARGQGVGPPGPKAAGEGPKGAGRGVPPGAVGPA 345
Db 372 GRTGPPGAGQDGRPPGPPGARGQGVGPPGPKAAGEGPKGAGRGVPPGAVGPA 431
Qy 346 GKDGAGAGQGGPPGPPGAGERGEGQGPAGSPGFGQLGPPGPPGPEAGKPCQGVPGDLGAP 405
Db 432 GKDGAGAGQGGPPGPPGAGERGEGQGPAGSPGFGQLGPPGPPGPEAGKPCQGVPGDLGAP 491
Qy 406 GPSGPAGE-----PGP-----TGLPDPGPPGRRGPPGRRFP 435
Db 492 GPSGARGRRGPPGRRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 551
Qy 436 GADGVAGPKGPPAGERGSPGAPGKSPGPE-----AGRGEAGLPGAKGLTSGSP 486
Db 552 GERGAAGLPKGRDGRDAGPKADGSPGKDGARGLTGPIGPPGPPGAPGDKGEAGSPGP 611
Qy 487 GPD-----GKTGPPGPPGAGQDGRPPGPPGARGQ-----AGVMGFP 522
Db 612 GPTGARGAPDGEAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 671
Qy 523 GPKGAAGPPGKAGRGVPPGPPG-----VGPAGKDGAGAGQPPGPPGPPGPP 570
Db 672 GPIGNVAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 731
Qy 571 GEQGPAGSPGFGQLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 616
Db 732 GETGPAGRGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 777

RESULT 10
US-08-468-996-10
Sequence 10, Application US/08468996
Patent No. 6645504
GENERAL INFORMATION:
APPLICANT: Weiner, Howard
APPLICANT: Miller, Ariel
APPLICANT: Zheng, Zheng
APPLICANT: Ahmad, Al-Sabbagh
TITLE OF INVENTION: BYSTANDER SUPPRESSION OF TYPE 1 DIABETES BY ORAL ADMINISTRATION OF
TITLE OF INVENTION: GLUCAGON
FILE REFERENCE: 1010/16959-US3
CURRENT APPLICATION NUMBER: US/08/468,996
CURRENT FILING DATE: 2003-02-07
PRIOR APPLICATION NUMBER: US 07/843,752
PRIOR FILING DATE: 1992-02-28
PRIOR APPLICATION NUMBER: US 07/460,852
PRIOR FILING DATE: 1990-02-21
PRIOR APPLICATION NUMBER: US 07/596,936
PRIOR FILING DATE: 1990-10-15
PRIOR APPLICATION NUMBER: US 07/065,734
PRIOR FILING DATE: 1987-06-24
PRIOR APPLICATION NUMBER: US 07/454,486
PRIOR FILING DATE: 1989-12-20
PRIOR APPLICATION NUMBER: US 07/487,732
PRIOR FILING DATE: 1990-03-02
PRIOR APPLICATION NUMBER: US 07/551,632
PRIOR FILING DATE: 1990-07-10
PRIOR APPLICATION NUMBER: US 07/379,778
PRIOR FILING DATE: 1989-07-14
PRIOR APPLICATION NUMBER: US 07/607,826
PRIOR FILING DATE: 1990-10-31
PRIOR APPLICATION NUMBER: US 07/595,468
NUMBER OF SEQ ID NOS: 13
SOFTWARE: Patent in version 3.1
SEQ ID NO 10
LENGTH: 1017
TYPE: PRT
ORGANISM: Homo sapiens
US-08-468-996-10

Query Match 56.0%; Score 1955; DB 4; Length 1017;
Best Local Similarity 52.5%; Pred. No. 5.9e-116;
Matches 396; Conservative 38; Mismatches 182; Indels 138; Gaps 15;

Qy 1 GPPGPPGPTGLPDPGPPGRRGPPGAGDGVAGPKGPPAGERGSPGPPGPKGSPGAGRRP 60
Db 25 GNPGEPPGVPSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 84
Qy 61 -----GEAGLPKAGLGTGSP-----GSPGP-----DGKTGPPGPA--- 90
Db 85 GVKHGRGVPGLDGAAGAGAPGVKSGSPGSPGSPGSPGSPGSPGSPGSPGSPGSPG 144
Qy 91 GQDGRPPGPPGPPGARGQAGVMGFP-----GPKGAAGPPGKAGRGVPPGPPGAVPA 141
Db 145 GNDQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 204
Qy 142 -----GKDGAGAGQGGPPGPPGAGERGEGQGPAGSPGFGQLGPPGPPGPEAGKPCQGV 195
Db 205 GASGNPCTDGTGPKAGSAGAPGAGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 264
Qy 196 GD---LGAPSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 237
Db 265 GEQGPKEPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 324
Qy 238 GPKGPPAGERGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 297
Db 325 GPKGPPAGERGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 384
Qy 298 GRPDPGPPGARGQGVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 357

PRIOR APPLICATION DATA: US/08/187,319
APPLICATION NUMBER: 21-JAN-1994
FILING DATE: 21-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN -ALPHA 1 (II)
US-08-963-825-20

Query Match 55.7%; Score 1943; DB 3; Length 1418;
Best Local Similarity 52.1%; Pred. No. 4.5e-115;
Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;
QY 1 GPPGPGTGLPFGPGRGSGRFGAGDVGAGPKGAGSGPAGPKGSGPGEAGRP 60
DB 156 GNPGEFPGVSGPMPGPPGPKGDDGEAGPKGAGSGPAGPKGSGPGEAGRP 215
QY 61 -----GBAGLPGAKGLTGP-----GSPGP-----DGKTPGPPGA--- 90
DB 216 GVKHGRGPGLDGAKGAGAGPVGKSGSPGNGSPGPMGRGLPGERGRTGPAGAAG 275
QY 91 GQDGRPPGPPGAGQAGVGMFP-----GPKGAAGPAGKAGRGVGPAGVGA 141
DB 276 GNDGQFAGPVPVGPAGGFPAGAPKAGKAGAGTGARGEGAGPGEFTGPGSPGA 335
QY 142 -----GKDEAGAAGPAGPAGGERGEGQFAGSGFGLPGLPAGPAGPAGKAGPGEAGV 195
DB 336 GASGNPDTGIPGAKGAGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 395
QY 196 GD-----LGAPSGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 237
DB 396 GQGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 455
QY 238 GPKGAGSGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 297
DB 456 GPKGAGSGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 515
QY 298 GRPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 357
DB 516 GRPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 575
QY 358 GPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 413
DB 576 GPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 635
QY 414 -----PGTGLP-----GPPGSGPAGPAGPAGPAGPAGPAGPAGPAG 447
DB 636 GERGSGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 695
QY 448 GERGS-----PGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 483
DB 696 GDRGVGKPPGAGPKDGRGLTGP-----GPPGAGPAGPAGPAGPAGPAG 755
QY 484 GSPGDPGKTPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 534
DB 756 GTGTPGTSGIAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 815
QY 535 GERGVGPPGA-----VCPAGKDEAGAGPAGPAGPAGPAGPAGPAGPAG 582

DB 816 GARGAQPPGATGPPGAAGRVGPPGSGNPGPPGSGKDGKPGKAGSDSGPPGRAGEP 875
QY 583 GLPAPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 616
DB 876 GLQAPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 909
RESULT 13
US-09-010-999-1
Sequence 1, Application US/09010999
Patent No. 6132976
GENERAL INFORMATION:
APPLICANT: Poole, Anthony R.
APPLICANT: Hollander, Anthony P.
APPLICANT: Billingham, R. C.
TITLE OF INVENTION: IMMUNOASSAYS FOR THE MEASUREMENT OF
TITLE OF INVENTION: COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,999
FILING DATE: 22-JAN-1998
CLASSIFICATION: 4335
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/448,501
FILING DATE: 17-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/984,123
FILING DATE: 04-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 032931/0212
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1418 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Human Type II Collagen
US-09-010-999-1
Query Match 55.7%; Score 1943; DB 3; Length 1418;
Best Local Similarity 52.1%; Pred. No. 4.5e-115;
Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;
QY 1 GPPGPGTGLPFGPGRGSGRFGAGDVGAGPKGAGSGPAGPKGSGPGEAGRP 60
DB 156 GNPGEFPGVSGPMPGPPGPKGDDGEAGPKGAGSGPAGPKGSGPGEAGRP 215
QY 61 -----GBAGLPGAKGLTGP-----GSPGP-----DGKTPGPPGA--- 90
DB 216 GVKHGRGPGLDGAKGAGAGPVGKSGSPGNGSPGPMGRGLPGERGRTGPAGAAG 275
QY 91 GQDGRPPGPPGAGQAGVGMFP-----GPKGAAGPAGKAGRGVGPAGVGA 141

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 15:02:42 ; Search time 67.5856 Seconds
(without alignments)
3505.413 Million cell updates/sec

Title: US-10-658-989A-2

Perfect score: 3488

Sequence: 1 GPPGEPGTGLPGPPGRRG.....GEOGVGDGLGPGSPGAGG 617

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1714042 seqs, 363979560 residues

Total number of hits satisfying chosen parameters: 1714042

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3488	100.0	617	17	US-10-658-989A-2
2	3482	99.8	821	17	US-10-658-989A-3
3	2237.5	64.1	1014	17	US-10-901-816A-5
4	2237.5	64.1	1014	17	US-10-901-816A-6
5	2237.5	64.1	1014	17	US-10-901-816A-10
6	2237.5	64.1	1057	15	US-10-104-889-20
7	2237.5	64.1	1107	15	US-10-104-889-11
8	2237.5	64.1	1171	15	US-10-104-889-8
9	2237.5	64.1	1388	15	US-10-104-889-10
10	2237.5	64.1	1464	16	US-10-216-705-21
11	2237.5	64.1	1464	16	US-10-357-851-1

12	2237.5	64.1	1464	16	US-10-358-024-1	Sequence 1, Appli
13	2237.5	64.1	1464	16	US-10-788-792-150	Sequence 150, App
14	2235.5	64.1	1014	17	US-10-901-816A-9	Sequence 9, Appli
15	2235.5	64.1	1014	17	US-10-901-816A-11	Sequence 11, Appli
16	2234	64.0	1449	15	US-10-402-089-8	Sequence 8, Appli
17	2234	64.0	1449	15	US-10-402-072A-8	Sequence 8, Appli
18	2233.5	64.0	1461	16	US-10-468-091-25	Sequence 25, Appli
19	2233.5	64.0	1464	15	US-10-291-265-243	Sequence 243, App
20	2231.5	64.0	1057	15	US-10-104-889-16	Sequence 16, Appli
21	2228.5	63.9	1341	14	US-10-058-124-18	Sequence 18, Appli
22	2227.5	63.9	1464	10	US-09-918-715-261	Sequence 261, App
23	2227.5	63.9	1464	14	US-10-060-036-159	Sequence 159, App
24	2227.5	63.9	1464	14	US-10-171-311-36	Sequence 36, Appli
25	2227.5	63.9	1464	14	US-10-149-352-2	Sequence 2, Appli
26	2227.5	63.9	1464	14	US-10-177-293-65	Sequence 55, Appli
27	2227.5	63.9	1464	14	US-10-301-822-28	Sequence 28, Appli
28	2227.5	63.9	1464	16	US-10-734-564-79	Sequence 79, Appli
29	2227.5	63.9	1464	16	US-10-474-794-261	Sequence 261, App
30	2227.5	63.9	1464	16	US-10-723-860-2289	Sequence 2289, Ap
31	2227.5	63.9	1464	17	US-10-852-335A-157	Sequence 157, App
32	2225.5	63.8	1463	15	US-10-402-089-2	Sequence 2, Appli
33	2225.5	63.8	1463	15	US-10-402-072A-2	Sequence 2, Appli
34	2224.5	63.8	1014	17	US-10-901-816A-7	Sequence 7, Appli
35	2224.5	63.8	1014	17	US-10-901-816A-13	Sequence 13, Appli
36	2222.5	63.7	1014	17	US-10-901-816A-8	Sequence 8, Appli
37	2222.5	63.7	1169	15	US-10-104-889-6	Sequence 6, Appli
38	2205.5	63.2	1453	16	US-10-468-091-26	Sequence 26, Appli
39	2193.5	62.9	1014	17	US-10-901-816A-12	Sequence 12, Appli
40	2191	62.8	822	15	US-10-342-331-49	Sequence 49, Appli
41	1955	56.0	1017	16	US-10-639-286-10	Sequence 10, Appli
42	1946	55.8	1014	14	US-10-194-441A-1	Sequence 1, Appli
43	1943	55.7	1418	14	US-10-058-124-20	Sequence 20, Appli
44	1943	55.7	1418	16	US-10-468-091-5	Sequence 5, Appli
45	1938	55.6	544	17	US-10-658-989A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-10-658-989A-2
; Sequence 2, Application US/10658989A
; Publication No. US20050101531A1

; GENERAL INFORMATION:

; APPLICANT: BOWSTRA, Jan Bastiaan

; APPLICANT: YUZO, Toda

; TITLE OF INVENTION: Use of recombinant gelatin-like proteins as plasma expanders and
; FILE REFERENCE: BOWSTRA-3

; CURRENT APPLICATION NUMBER: US/10/658,989A

; CURRENT FILING DATE: 2003-09-10

; PRIOR FILING DATE: 2002-09-11

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 617

; TYPE: PRT

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Hu-3

US-10-658-989A-2

Query Match	100.0%;	Score 3488;	DB 17;	Length 617;
Best Local Similarity	100.0%;	Pred. No. 3.9e-167;		
Matches 617;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GPPGEPGTGLPGPPGRRGSGRFGADGAGKPGAGRGSPGAGKSPGACRGP	60	
Db	1	GPPGEPGTGLPGPPGRRGSGRFGADGAGKPGAGRGSPGAGKSPGACRGP	60	
Qy	61	GEAGLPGAKGLTSGPSGPPGDKTGPAGODRPPGPPGARGQGVGFGPKGAA	120	

Db 61 GEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRPPGPPGARGQAGVMGFGPKGAA 120
 QY 121 GEPGKAGRGVPGPPGAVGAPAGKDCGAGAGQAGPPGAPAGRGSGQAGSGPFGQGLPGPA 180
 Db 121 GEPGKAGRGVPGPPGAVGAPAGKDCGAGAGQAGPPGAPAGRGSGQAGSGPFGQGLPGPA 180
 QY 181 GPPGKAGRGVPGPPGAVGAPAGKDCGAGAGQAGPPGAPAGRGSGQAGSGPFGQGLPGPA 240
 Db 181 GPPGKAGRGVPGPPGAVGAPAGKDCGAGAGQAGPPGAPAGRGSGQAGSGPFGQGLPGPA 240
 QY 241 GPAGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 300
 Db 241 GPAGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 300
 QY 301 GPPGPPGARGQAGVMGFGPKGAAAGFPKAGRGVPGPPGAVGAPAGKDCGAGAGQAGPPGPA 360
 Db 301 GPPGPPGARGQAGVMGFGPKGAAAGFPKAGRGVPGPPGAVGAPAGKDCGAGAGQAGPPGPA 360
 QY 361 GPAGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 420
 Db 361 GPAGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 420
 QY 421 GPPGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 480
 Db 421 GPPGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 480
 QY 481 GSPGSGPDKTGPAGQDGRP 540
 Db 481 GSPGSGPDKTGPAGQDGRP 540
 QY 541 GPPGAVGAPAGKDCGAGAGQAGPPGAPAGRGSGQAGSGPFGQGLPGPA 600
 Db 541 GPPGAVGAPAGKDCGAGAGQAGPPGAPAGRGSGQAGSGPFGQGLPGPA 600
 QY 601 GVPDGLGAPGSGPAGG 617
 Db 601 GVPDGLGAPGSGPAGG 617

RESULT 2
 US-10-658-989A-3
 ; Sequence 3, Application US/10658989A
 ; Publication No. US2005010151A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BOWSTRA, Jan Bastiaan
 ; APPLICANT: YUZO, Toda
 ; TITLE OF INVENTION: Use of recombinant gelatin-like proteins as plasma expanders and
 ; FILE REFERENCE: compositions suitable for plasma substitution
 ; CURRENT APPLICATION NUMBER: US/10/658,989A
 ; PRIOR FILING DATE: 2003-09-10
 ; PRIOR FILING DATE: 2002-09-11
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: Patent in version 3.1
 ; SEQ ID NO 3
 ; LENGTH: 821
 ; TYPE: PRT
 ; ORGANISM: Artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: Hu-4
 US-10-658-989A-3

Query Match 99.8%; Score 3482; DB 17; Length 821;
 Best Local Similarity 100.0%; Pred. No. 9.5e-167;
 Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GPPGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 60
 Db 1 GPPGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 60
 QY 61 GEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 120
 Db 61 GEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 120

Db 61 GEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 120
 QY 121 GEPGKAGRGVPGPPGAVGAPAGKDCGAGAGQAGPPGAPAGRGSGQAGSGPFGQGLPGPA 180
 Db 121 GEPGKAGRGVPGPPGAVGAPAGKDCGAGAGQAGPPGAPAGRGSGQAGSGPFGQGLPGPA 180
 QY 181 GPPGKAGRGVPGPPGAVGAPAGKDCGAGAGQAGPPGAPAGRGSGQAGSGPFGQGLPGPA 240
 Db 181 GPPGKAGRGVPGPPGAVGAPAGKDCGAGAGQAGPPGAPAGRGSGQAGSGPFGQGLPGPA 240
 QY 241 GPAGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 300
 Db 241 GPAGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 300
 QY 301 GPPGPPGARGQAGVMGFGPKGAAAGFPKAGRGVPGPPGAVGAPAGKDCGAGAGQAGPPGPA 360
 Db 301 GPPGPPGARGQAGVMGFGPKGAAAGFPKAGRGVPGPPGAVGAPAGKDCGAGAGQAGPPGPA 360
 QY 361 GPAGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 420
 Db 361 GPAGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 420
 QY 421 GPPGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 480
 Db 421 GPPGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 480
 QY 481 GSPGSGPDKTGPAGQDGRP 540
 Db 481 GSPGSGPDKTGPAGQDGRP 540
 QY 541 GPPGAVGAPAGKDCGAGAGQAGPPGAPAGRGSGQAGSGPFGQGLPGPA 600
 Db 541 GPPGAVGAPAGKDCGAGAGQAGPPGAPAGRGSGQAGSGPFGQGLPGPA 600
 QY 601 GVPDGLGAPGSGPAG 616
 Db 601 GVPDGLGAPGSGPAG 616

RESULT 3
 US-10-901-816A-5
 ; Sequence 5, Application US/10901816A
 ; Publication No. US20050059703A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Chang, Robert C.
 ; APPLICANT: Olsen, David R.
 ; APPLICANT: James, Polarek W.
 ; APPLICANT: Williams, Kim E.
 ; TITLE OF INVENTION: Gelatin Capsules
 ; FILE REFERENCE: PP0404 US
 ; CURRENT APPLICATION NUMBER: US/10/901,816A
 ; PRIOR FILING DATE: 2004-07-29
 ; PRIOR FILING DATE: 2003-08-01
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: Patent in version 3.2
 ; SEQ ID NO 5
 ; LENGTH: 1014
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-901-816A-5

Query Match 64.1%; Score 2237.5; DB 17; Length 1014;
 Best Local Similarity 60.2%; Pred. No. 1.7e-104;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;
 QY 1 GPP-----CEPPTGLPGPPGERSGPGAGPAGQDGRP 48
 Db 280 GPPGERSGPGAPGKSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 48
 QY 49 GPKGSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 108
 Db 340 GPKGSGPGEAGLPGAKGLTSGSPGSGPDKTGPAGQDGRP 108

; ORGANISM: Homo sapiens
US-10-901-816A-10

Query Match 64.1%; Score 2237.5; DB 17; Length 1014;
Best Local Similarity 60.2%; Pred. No. 1.7e-104;
Matches 441; Conservative 23; Mismatches 152; Indels 117;

QY	1	GPP	-----GPPGPTGLPGPPGERGGPSRGGPGADGVAGPKPGAGERGSPGA	48
Db	280	GPFPAGEGKRGARGEFGPTGLPPTGERGGPSRGGPGADGVAGPKPGAGERGSPGA	339	
QY	49	GPXGSPGEACRPGEBAGLPGAKGLTGSFGSPGPDGKTGPPAGQDGRPPGPPGARGQA	108	
Db	340	GPXGSPGEACRPGEBAGLPGAKGLTGSFGSPGPDGKTGPPAGQDGRPPGPPGARGQA	399	
QY	109	GVMGFFCPKKAAGBPCKAGRGVPPGAVGVPACKDGEAGAQQGPPGAPGABRGEGOGPA	168	
Db	400	GVMGFFCPKKAAGBPCKAGRGVPPGAVGVPACKDGEAGAQQGPPGAPGABRGEGOGPA	459	
QY	169	GSFGFQGLPGAPPPGABGPEQGVPGDLGAPGSPGAGE	212	
Db	460	GSFGFQGLPGAPPPGABGPEQGVPGDLGAPGSPGABRGPPGPPGPPGAPGR	519	
QY	213	-----TGLPBPBERGGPSRGFPDGVAGPKGABRGSGPSGAPGPKGSP	258	
Db	520	GANGAPNDGAKGDAGAPGAGSQGAPGLOQMPGERGAAGLPGPKGDRGDAGPKGADGSP	579	
QY	259	GE-----AGRPGBAGLPGAK-----GLTGSFGSPGPDGKTGPPGAGQDGRP	300	
Db	580	GKDGVRLTGPIGPPGPPGAPGDKGESGSPGAPGTGARGAPGDRGPPGPPGAPGAP	639	
QY	301	GPFPGPPGARGQ-----AGVMGFFCPKKAAGBPCKAGRGVGGPPGA	341	
Db	640	GADGQPCAKGEPGDAGAKGDAGPPGAPGAPGPIGPPCAPKAGARGSAGSPGATGFP	699	
QY	342	-----VGPAGKDEAGNQGPPGAPG---AGERGEQGPAGSPGQGLPGABGPPGABGKP	393	
Db	700	GAAGRVPGPSPGNAGPPGPPGABGEGKGRGETGTPARCPGEVPPGPPGABGKGP	759	
QY	394	GEQGVPGDLGAPSPGABPGPTGLPGPPGERGGPSRGGFPDGVAGPKGABRGSP	453	
Db	760	GADGPAGATPTPQGLIAGRGVVUPLPGQKRGFFPLPGSPGPKQGPSGASGSRGPP	819	
QY	454	GPAGP---KGSPEACRPGEBAGLPGAKGLTGSFGSPGPDGKTGPPAGQDGRPPGPPGP	510	
Db	820	GPWGPPLGAPPPGESCREGAPGAEGSPRGDGSFGAKGDRCEITGPAGPCGAPGAPGPV	879	
QY	511	GARGQAVMGFPKPAAGEPGKABRGVPPGAVGVPAGKQGE	558	
Db	880	GPAGKSGDRGETGAPGAPGVGAPGARGPAGPQGRDGTGEQDGRGDKHGRGFSGLQ	939	
QY	559	GPFGP-----AGPABRGQGPAGS---GFGCLPGPAGPPGEBAGKPGGQGP	603	
Db	940	CTPFGPPSGPEQSGCASGFAPGRPPGGSNAGPKDGLNGLPLGPIGPPGRGTGDAGPV	999	
QY	604	GDLGAFGPGGAPG	616	
Db	1000	GPFPGPPGPPG	1012	

RESULT 6

US-10-104-889-20.
; Sequence 20, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:

APPLICANT: GRUSKIN, ELLIOT A.
BUECHTER, DOUGLAS
BROKAW, JANE
ZHANG, GUANGHUI
PAOLELLA, DAVID

PROLETA, DAVID
 TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:

```

1 / ADDRESSEE: DILLWORTH & BARRESE
2 / STREET: 333 EARLE OVINGTON BOULEVARD
3 / CITY: UNIONDALE
4 / STATE: NY
5 / COUNTRY: U.S.A.
6 / ZIP: 11553
7 /
8 / COMPUTER READABLE FORM:
9 / MEDIUM TYPE: Floppy disk
10 / COMPUTER: IBM PC compatible
11 / OPERATING SYSTEM: PC-DOS/MS-DOS
12 / SOFTWARE: Patent In Release #1.0, Ver
13 / CURRENT APPLICATION DATA:
14 / APPLICATION NUMBER: US/10/104,889
15 / FILING DATE: 22-Mar-2002
16 / CLASSIFICATION: <Unknown>
17 / PRIOR APPLICATION DATA:
18 / APPLICATION NUMBER: US/09/169,768
19 / FILING DATE: 09-OCT-1998
20 / ATTORNEY/AGENT INFORMATION:
21 / NAME: STEEN, JEFFREY S
22 / TELECOMMUNICATION INFORMATION:
23 / TELEPHONE: (516) 228-8484
24 / TELEFAX: (516) 228-8516
25 / INFORMATION FOR SEQ ID NO: 20:
26 / SEQUENCE CHARACTERISTICS:
27 / LENGTH: 1057 amino acids
28 / TYPE: amino acid
29 / STRANDEDNESS: single
30 / TOPOLOGY: unknown
31 / MOLECULE TYPE: peptide
32 / SEQUENCE DESCRIPTION: SEQ ID NO: 20:
33 / US-10-104-889-20

```

Query Match	64.1%; Score 2237.5; DB 15; Length 1057;
Best Local Similarity	60.2%; Pred. No. 1.7e-104;
Matches	441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;
QY	1 GPP-----CEPGTGLPGPPGRRGGGSGRFPFGADGVAGPKGPAGERGSGPPA 48
Db	
297	GPGPAGECKRGARBPFTGLPGPPGRRGGGSGRFPFGADGVAGPKGPAGERGSGPPA 356
QY	49 GPXGSGEAGRPGAEGLPGAKGLTGSPCSPDPGKTGPPGAGDGRPPGPPPGAGQA 108
Db	
357	GPXGSGEAGRPGAEGLPGAKGLTGSPCSPDPGKTGPPGAGDGRPPGPPPGAGQA 416
QY	109 GVMGFPKPKAAECPKAGRGVPPGPAVGPAGKDGCAQAQGGPPGAGPAGERGEGQPA 168
Db	
417	GVMGFPKPKAAECPKAGRGVPPGPAVGPAGKDGCAQAQGGPPGAGPAGERGEGQPA 476
QY	169 GSPFGGLPGPAGPPCEAGKPGQGVPGDLGACQSPSGAG-----PGP----- 212
Db	
477	GSPFGGLPGPAGPPCEAGKPGQGVPGDLGACQSPSGAG-----PGP----- 536
QY	213 -----TGLPGPPGRRGGGSGRFPFGADGVAGPKGPAGERGSGPPGAPKGPSP 258
Db	
537	GANGAPNDGAKGDACAGPAGSQAGPGLQGMFGERGAAGLPKPGDRGDAGPKAGDGPSP 596
QY	259 GE-----AGRPGAEGLPGAK-----GLTGSPPSGPDGKTGPPGAPQDGRP 300
Db	
597	GKDGVRGLTGPIPPPGPAGPDGKBSGSPSGPAGPTGARGAPDGRGPPGPPGAGFAGPP 656
QY	301 GPPGPPGARGO-----AGVMGFPKPKAAECPKAGRGVPPGPPCA----- 341
Db	
657	GADGQCPAKGEPDGAAGDAGPPGAPGAPPGPIGNVGPAGKARGAGSGPPGATGFP 716
QY	342 -----VGPAGKDGAEAGAPPPGPPAGP---AGERGEQGPAGSPFGILPGPAGPPGEAGKP 393
Db	
717	GAAGRVPDPGPGSNAGPPGPPGAGKEGKGPRGETGPAGRPGVEVPPGPPGAGEKGPSP 776
QY	394 GSGQGVPCDILGAPGSPGAGEPPTGLPGPPGRRGGGSGRFPFGADGVAGPKGPAGERGSP 453
Db	
777	GADGPAGAPCTGPPQGIAGQRVGVGLPGQRRGERGFFLPGSPGEPKQKQSGSAGSGRPP 836


```

; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1171 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-104-889-8

```

```

Query Match      64.1%; Score 2237.5; DB 15; Length 1171;
Best Local Similarity 60.2%; Pred. No. 1.8e-104;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

Qy 1 GPP-----GPPGPTGLPGRGSGRGGFPAGDGVAGPKPAGERGSPGPA 48
Db 297 GPPGAGEGKRGARGEPGPTGLPGRGSGRGGFPAGDGVAGPKPAGERGSPGPA 356

Qy 49 GPKGSGEAGRPGAEGLPGAKGLTGSFGSPGPGKTPGPPAGQDGRPPGPPGARGQA 108
Db 357 GPKGSGEAGRPGAEGLPGAKGLTGSFGSPGPGKTPGPPAGQDGRPPGPPGARGQA 416

Qy 109 GVMGFPKGAAGHPKAGRGVPPGPGAVGPGAGKDGAGAGQPPGPPGAGERGQGA 168
Db 417 GVMGFPKGAAGHPKAGRGVPPGPGAVGPGAGKDGAGAGQPPGPPGAGERGQGA 476

Qy 169 GSPGFQGLPGAPGPPGAEAGKPGEGVPGDLGAPGSGPAGE-----PGP----- 212
Db 477 GSPGFQGLPGAPGPPGAEAGKPGEGVPGDLGAPGSGPAGE-----PGP----- 536

Qy 213 -----TGLPGRGSGRGGFPAGDGVAGPKPAGERGSPGPA 258
Db 537 GANAPGNDGAKGDAGAPGAGPGLQGMPSGGERGAAGLPKPGDGRDAGPKADGSP 596

Qy 259 GE-----ACRPGFAGLPGAK-----GLTGSFGSPGPGKTPGPPAGQDGRPP 300
Db 597 GKDGVRLGTPIPPGPGAGPKGSGSPGAGPTGARGAPGDRGEPGPPGAGFAGPP 656

Qy 301 GPPGPPGARGQ-----AGVMGFPKGAAGHPKAGRGVPPGPGAVGPGAGKDGAG 341
Db 657 GADGQPGAKGPECDAGAKGDAGPPGAGPAGPPIGNVGAFCAGKAGSAGPAGTATFP 716

Qy 342 -----VGPAGKDGAEAGQPPGAGP-----AGRGGSGGAGSPGFGGLPGAGPPGAEAGK 393
Db 717 GAAGRVGPPGSGNAGPPGPPGAGKEGKGRGETGAGRPGEVGPPGPPGAGEKSP 776

Qy 394 GEQGVPGDLGAPGSPGAPGPTGLPGRGSGRGGFPAGDGVAGPKPAGERGSP 453
Db 777 GADGPAGAGTGTGQIACQGVVGLPQGRGGRGFPGLPGSGEPGKPGSGASGERGPP 836

Qy 454 GPAGP---KGSPGAEAGRPGAEGLPGAKGLTGSFGSPGPGDGTGPPGAGQDGRPPGPP 510
Db 837 GPMGPPGLAGPPGSGREGAPGAGSGPGRDGSFAGKDGRTGTGAPGPPGAGPAGPVP 896

Qy 511 GARGQAVMGFPKGAAGHPKAGRGVPPGPGAVGPGAGKDG-----AGAQ 558
Db 897 GPAGKSGDRGTGTGAPGAPGVPVGAARGPAGPQGPGRGDKGETGEGQDGRGKGRGFSGLQ 956

Qy 559 GPPGP-----AGPAGERGQPPAGSP-----GFGELPGPAGPPGAEAGKPGEGVP 603
Db 957 GPPGPPGSGEQSPGASGAPGPPGASAGAPKDGKGLNGLPGLPGRGRTGDAGVP 1016

Qy 604 GDLGAPGSGPAG 616

```

```

Db 1017 GPPGPPGPPGPPG 1029

RESULT 9
US-10-104-889-10
; Sequence 10, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: BRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVERTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-104-889-10

```

```

Query Match      64.1%; Score 2237.5; DB 15; Length 1388;
Best Local Similarity 60.2%; Pred. No. 2.1e-104;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

Qy 1 GPP-----GPPGPTGLPGRGSGRGGFPAGDGVAGPKPAGERGSPGPA 48
Db 297 GPPGAGEGKRGARGEPGPTGLPGRGSGRGGFPAGDGVAGPKPAGERGSPGPA 356

Qy 49 GPKGSGEAGRPGAEGLPGAKGLTGSFGSPGPGKTPGPPAGQDGRPPGPPGARGQA 108
Db 357 GPKGSGEAGRPGAEGLPGAKGLTGSFGSPGPGKTPGPPAGQDGRPPGPPGARGQA 416

Qy 109 GVMGFPKGAAGHPKAGRGVPPGPGAVGPGAGKDGAGAGQPPGPPGAGERGQGA 168
Db 417 GVMGFPKGAAGHPKAGRGVPPGPGAVGPGAGKDGAGAGQPPGPPGAGERGQGA 476

Qy 169 GSPGFQGLPGAPGPPGAEAGKPGEGVPGDLGAPGSGPAGE-----PGP----- 212
Db 477 GSPGFQGLPGAPGPPGAEAGKPGEGVPGDLGAPGSGPAGE-----PGP----- 536

Qy 213 -----TGLPGRGSGRGGFPAGDGVAGPKPAGERGSPGPA 258
Db 537 GANAPGNDGAKGDAGAPGAGPGLQGMPSGGERGAAGLPKPGDGRDAGPKADGSP 596

```


QY 259 GE-----AGRGEAGLPGAK-----GLTSPSPGPDGKTGTPPGPAGQDGRP 300
 Db 597 GKGVRLGTGIPGPPAGAFDKGESGSPGAPGTGARGAPGDRGEPGPPGAPGAGPP 656
 QY 301 GPPGPPGARGO-----AGVMGFPKGAAGFPKAGRGVGPFGCA----- 341
 Db 657 GADQOPKAGPFGDAGAKDAGPPGAPGAPGPPGPIGNVGAFAKAGSAGPFGATGFP 716
 QY 342 -----VGPAGKDBAGAGQPPGAPG---AGERGEQGPAGSPGFGQLPGPAGPPGEGAGKP 393
 Db 717 GAAGRVGPPGSGNAGPPGPPGAGKEGKGRGTGAPRGVEGVPDPGPPGAGEXGSP 776
 QY 394 GEQVPGDLGAPGSPGAPGPTGLPCPPPERGPGSGRPGADGVAGPKPAGERGSP 453
 Db 777 GADGPAGAGTTPGQGTAGQGVVGLPGQORGERFPGLPGSGFPGKQSGSGASGERGPP 836
 QY 454 GPAGP---KGSPPGABRPFGEAGLPGAKGLTSGSPSGPDGKTGTPPGPAGQDGRP 510
 Db 837 GPMGPPGLAGPPGSGREGAFGAGSGPDRDGSFGAKGDRGTGAPGPPGAPGAPGVP 896
 QY 511 GARGQAGVMGPPGPKGAAGEFGKAGRGVPPGPAVGPAGKDG-----AGAQ 558
 Db 897 GPAGKSGDRGTGAPGAPGVPAGARGPAGPQGPGRGDKGTGEGQDRGKIGHRGFSGLQ 956
 QY 559 GPPGP-----AGPAGERGEQGPAGSP---GFGQLPGPAGPPGEGAKGPEQGV 603
 Db 957 GPPGPPGSGPQSPGASGAPGPPGSGAGAPGKQGLNGLPPIGPPGPRGRTGDAGPV 1016
 QY 604 GDLGAPGSPGAPG 616
 Db 1017 GPPGPPGPPGPPG 1029

RESULT 10
 US-10-216-705-21
 ; Sequence 21, Application US/10216705
 ; Publication No. US20030096973A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Meristem Therapeutics, S.A.
 ; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Me
 ; TITLE OF INVENTION: obtaining Such and Their Uses
 ; FILE REFERENCE: 1149-3 DIV
 ; CURRENT APPLICATION NUMBER: US/10/216,705
 ; CURRENT FILING DATE: 2002-08-09
 ; PRIOR APPLICATION NUMBER: US 09/331,347
 ; PRIOR FILING DATE: 1999-08-17
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 21
 ; LENGTH: 1464
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-216-705-21

Query Match 64.1%; Score 2237.5; DB 14; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 2.2e-104;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPCPPGERGPGSGRPGADGVAGPKPAGERGSPGPA 48
 Db 458 GPPGAGEGKRGARGEFGPTGLPGPPGERGPGSGRPGADGVAGPKPAGERGSPGPA 517
 QY 49 GPKSGPAGRPFGEAGLPGAKGLTSGSPSGPDGKTGTPPGAGQDGRP 108
 Db 518 GPKSGPAGRPFGEAGLPGAKGLTSGSPSGPDGKTGTPPGAGQDGRP 577
 QY 109 GVMGPPGPKGAAGEFGKAGRGVPPGPAVGPAGKDGAGAGQPPGAPGAGERGEGQPA 168
 Db 578 GVMGPPGPKGAAGEFGKAGRGVPPGPAVGPAGKDGAGAGQPPGAPGAGERGEGQPA 637
 QY 169 GSPGFOGLPGPAGPPGEGAKGPEQGVPGDILGAPGSPGAGE-----PGP----- 212

Db 638 GSPGFOGLPGPAGPPGEGAKGPEQGVPGDILGAPGSPGARGERGPPGVPQPPGAPGR 697
 QY 213 -----TGLPGPPGERGPGSGRPGADGVAGPKPAGERGSPGAPGKGPSP 258
 Db 698 GAGAPGNDGAKGADAGAPGAPGSGQAFGLQMPGERGAAGLPGPKGDRGDAGPAGDAGSP 757
 QY 259 GE-----AGRPFGEAGLPGAK-----GLTSPSPSGPDGKTGTPPGPAGQDGRP 300
 Db 758 GKGVRLGTGIPGPPGAPGAPGDKGESGSPGAPGTGARGAPGDRGEPGPPGAPGAGPP 817
 QY 301 GPPGPPGARGO-----AGVMGFPKGAAGFPKAGRGVGPFGCA----- 341
 Db 818 GADQOPKAGPFGDAGAKDAGPPGAPGAPGPPGPIGNVGAFAKAGSAGPFGATGFP 877
 QY 342 -----VGPAGKDBAGAGQPPGAPG---AGERGEQGPAGSPGFGQLPGPAGPPGEGAGKP 393
 Db 878 GAAGRVGPPGSGNAGPPGPPGAGKEGKGRGTGAPRGVEGVPDPGPPGAGEXGSP 937
 QY 394 GEQVPGDLGAPGSPGAPGPTGLPCPPPERGPGSGRPGADGVAGPKPAGERGSP 453
 Db 938 GADGPAGAGTTPGQGTAGQGVVGLPGQORGERFPGLPGSGFPGKQSGSGASGERGPP 997
 QY 454 GPAGP---KGSPPGABRPFGEAGLPGAKGLTSGSPSGPDGKTGTPPGPAGQDGRP 510
 Db 998 GPMGPPGLAGPPGSGREGAFGAGSGPDRDGSFGAKGDRGTGAPGPPGAPGAPGVP 1057
 QY 511 GARGQAGVMGPPGPKGAAGEFGKAGRGVPPGPAVGPAGKDG-----AGAQ 558
 Db 1058 GPAGKSGDRGTGAPGAPGVPAGARGPAGPQGPGRGDKGTGEGQDRGKIGHRGFSGLQ 1117
 QY 559 GPPGP-----AGPAGERGEQGPAGSP---GFGQLPGPAGPPGEGAKGPEQGV 603
 Db 1118 GPPGPPGSGPQSPGASGAPGPPGSGAGAPGKQGLNGLPPIGPPGPRGRTGDAGPV 1177
 QY 604 GDLGAPGSPGAPG 616
 Db 1178 GPPGPPGPPGPPG 1190

RESULT 11
 US-10-357-851-1
 ; Sequence 1, Application US/10357851
 ; Publication No. US20040151731A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jicha, Douglas L.
 ; TITLE OF INVENTION: Method and Compositions Involving
 ; TITLE OF INVENTION: T-Lymphocyte Reactivity with Collagen
 ; TITLE OF INVENTION: Abdominal Aortic Aneurysm Patients
 ; FILE REFERENCE: 13376US
 ; CURRENT APPLICATION NUMBER: US/10/357,851
 ; CURRENT FILING DATE: 2003-02-04
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1
 ; LENGTH: 1464
 ; TYPE: PRT
 ; ORGANISM: Homo Sapiens
 US-10-357-851-1

Query Match 64.1%; Score 2237.5; DB 16; Length 1464;
 Best Local Similarity 60.2%; Pred. No. 2.2e-104;
 Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPCPPGERGPGSGRPGADGVAGPKPAGERGSPGPA 48
 Db 458 GPPGAGEGKRGARGEFGPTGLPGPPGERGPGSGRPGADGVAGPKPAGERGSPGPA 517
 QY 49 GPKSGPAGRPFGEAGLPGAKGLTSGSPSGPDGKTGTPPGAGQDGRP 108
 Db 518 GPKSGPAGRPFGEAGLPGAKGLTSGSPSGPDGKTGTPPGAGQDGRP 577
 QY 109 GVMGPPGPKGAAGEFGKAGRGVPPGPAVGPAGKDGAGAGQPPGAPGAGERGEGQPA 168
 Db 518 GVMGPPGPKGAAGEFGKAGRGVPPGPAVGPAGKDGAGAGQPPGAPGAGERGEGQPA 577
 QY 109 GVMGPPGPKGAAGEFGKAGRGVPPGPAVGPAGKDGAGAGQPPGAPGAGERGEGQPA 168

QY 1 GPP-----GEPGPTGLPGRGCGSRGPGADGVAGPKGAGERSGPGA 48
Db |||||
QY 458 GPPGAGEGKRGARGEPGPTGLPFPGRGCGSRGPGADGVAGPKGAGERSGPGA 517
Db |||||
QY 49 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGPPGARGQA 108
Db |||||
QY 518 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGARGQA 577
Db |||||
QY 109 GVMGFPKGAAGPCKAGRGVPPGCAVGPAGKDGEGAGAQGGPPGAGPAGERGEGQA 168
Db |||||
QY 578 GVMGFPKGAAGPCKAGRGVPPGCAVGPAGKDGEGAGAQGGPPGAGPAGERGEGQA 637
Db |||||
QY 169 GSPGFQGLPAGPAGPAGKRGEGVPGDLGAPGSPGAGE-----PGP----- 212
Db |||||
QY 638 GSPGFQGLPAGPAGPAGKRGEGVPGDLGAPGSPGARGERGPPGPPGPPGAPGR 697
Db |||||
QY 213 -----TGLPFPGRGCGSRGPGADGVAGPKGAGERSGPPGAPKGGSP 258
Db |||||
QY 698 GANCAPGNDGAKGDAGAPGAGPSQAGPCLQMPGERGAGLPGPKGDRGADGPKGADGSP 757
Db |||||
QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGPDGKTGPPGAGQDGRP 300
Db |||||
QY 758 GKQGVRLGTGPIGPPGAPAGPDKGSGSPGAGTGCARGAPGDRGPPGPPGAPGAGPP 817
Db |||||
QY 301 GPPGPPGARGO-----AGVMGFPCKGAAGEPKKAGRGVPPGPGA----- 341
Db |||||
QY 818 GADQCPGAKGEPGDAGAKGDAGPPGAPGAPGPPGPIGNVGAAPGAKGARGSAGSPGATGFP 877
Db |||||
QY 342 -----VGPAGKDGAGAQGGPPGAPG-----AGERGEOGPPGSPGQGLPAGPPGEGAGKP 393
Db |||||
QY 878 GAAGRVGPPGSPGNAGPPGPPGPPGAPKGGKGRGETGPPAGRGVEGPPGPPGAGEKSP 937
Db |||||
QY 394 GEQGVPGDLGAPGSPGAGEPPTGLPFPGRGCGSRGPPGADGVAGPKGAGERSGPP 453
Db |||||
QY 938 GADGAPAGAGTTPGQGIAGRGVVGLPGQRCGERGFPGLPGSPGPPGKQSPGASGERGPP 997
Db |||||
QY 454 GPAGP-----KGSPEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPP 510
Db |||||
QY 998 GPMGPPGLAGPPGSGREGAPGAEPSGRDGSFGAKGDRGETGPPGPPGAPGAPGCPV 1057
Db |||||
QY 511 GARGQAGVMGFPCKGAAGEPKKAGRGVPPGCAVGPAGKDG-----AGAQ 558
Db |||||
QY 1058 GPAGKSGDRGETGAPGAPGVPVGPAGAPGAPGQPRGDKGTGQDGRGKGRGFSGLQ 1117
Db |||||
QY 559 GPPGP-----AGPAGERGEOGPPGSP-----GFGQLPAGPPGPPGAGEKPPGQV 603
Db |||||
QY 1118 GPPGPPGSPGQSPGASGAPGPPGSGAGAPGKDGKGLNGLPGPIGPPGPRGRTGDAGPV 1177
Db |||||

RESULT 14

US-10-901-816A-9
; Sequence 9, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1014
; TYPE: PRT

; ORGANISM: Homo sapiens
US-10-901-816A-9

Query Match 64.1%; Score 2235.5; DB 17; Length 1014;
Best Local Similarity 59.9%; Pred. No. 2.1e-104;
Matches 439; Conservative 24; Mismatches 153; Indels 117; Gaps 11;

QY 1 GPP-----GEPGPTGLPGRGCGSRGPGADGVAGPKGAGERSGPGA 48
Db |||||
QY 280 GPPGAGEGKRGARGEPGPTGLPFPGRGCGSRGPGADGVAGPKGAGERSGPGA 339
Db |||||
QY 49 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGPPGARGQA 108
Db |||||
QY 340 GPKSGPEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGARGQA 399
Db |||||
QY 109 GVMGFPKGAAGPCKAGRGVPPGCAVGPAGKDGEGAGAQGGPPGAGPAGERGEGQA 168
Db |||||
QY 400 GVMGFPKGAAGPCKAGRGVPPGCAVGPAGKDGEGAGAQGGPPGAGPAGERGEGQA 459
Db |||||
QY 169 GSPGFQGLPAGPAGPAGKRGEGVPGDLGAPGSPGAGE-----PGP----- 212
Db |||||
QY 460 GSPGFQGLPAGPAGPAGKRGEGVPGDLGAPGSPGARGERGPPGPPGPPGAPGR 519
Db |||||
QY 213 -----TGLPFPGRGCGSRGPGADGVAGPKGAGERSGPPGAPKGGSP 258
Db |||||
QY 520 GANCAPGNDGAKGDAGAPGAGPSQAGPCLQMPGERGAGLPGPKGDRGADGPKGADGSP 579
Db |||||
QY 259 GE-----AGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRP 300
Db |||||
QY 580 GKQGVRLGTGPIGPPGAPAGPDPGSGSPGAGTGCARGAPGDRGPPGPPGAPGAGPP 639
Db |||||
QY 301 GPPGPPGARGO-----AGVMGFPCKGAAGEPKKAGRGVPPGPGA----- 341
Db |||||
QY 640 GADQCPGAKGEPGDAGAKGDAGPPGAPGAPGPPGPIGNVGAAPGAKGARGSAGSPGATGFP 699
Db |||||
QY 342 -----VGPAGKDGAGAQGGPPGAPG-----AGERGEOGPPGSPGQGLPAGPPGEGAGKP 393
Db |||||
QY 700 GAAGRVGPPGSPGNAGPPGPPGPPGAPKGGKGRGETGPPAGRGVEGPPGPPGAGEKSP 759
Db |||||
QY 394 GEQGVPGDLGAPGSPGAGEPPTGLPFPGRGCGSRGPPGADGVAGPKGAGERSGPP 453
Db |||||
QY 760 GADGAPAGACTTPGQGIAGRGVVGLPGQRCGERGFPGLPGSPGPPGKQSPGASGERGPP 819
Db |||||
QY 454 GPAGP-----KGSPEAGRPGEAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRPPGPP 510
Db |||||
QY 820 GPMGPPGLAGPPGSGREGAPGAEPSGRDGSFGAKGDRGETGPPGPPGAPGAPGCPV 879
Db |||||
QY 511 GARGQAGVMGFPCKGAAGEPKKAGRGVPPGCAVGPAGKDG-----AGAQ 558
Db |||||
QY 880 GPAGKSGDRGETGAPGAPGVPVGPAGAPGAPGQPRGDKGTGQDGRGKGRGFSGLQ 939
Db |||||
QY 559 GPPGP-----AGPAGERGEOGPPGSP-----GFGQLPAGPPGPPGAGEKPPGQV 603
Db |||||
QY 940 GPPGPPGSPGQSPGASGAPGPPGSGAGAPGKDGKGLNGLPGPIGPPGPRGRTGDAGPV 999
Db |||||
QY 604 GDLGAPGSPGAP 616
Db |||||
QY 1000 GPPGPPGPPGPPG 1012
Db |||||

RESULT 15

US-10-901-816A-11
; Sequence 11, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29

; PRIOR APPLICATION NUMBER: US 60/492,085
 ; PRIOR FILING DATE: 2003-08-01
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: Patent in version 3.2
 ; SEQ ID NO 11
 ; LENGTH: 1014
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-901-816A-11

Query Match 64.1%; Score 2235.5; DB 17; Length 1014;
 Best Local Similarity 59.9%; Pred. No. 2.1e-104;
 Matches 439; Conservative 24; Mismatches 153; Indels 117; Gaps 11;

Qy	1	GPP-----GPPGTGLPPPPGSGSGRFPADGVAGPKGAGERSFGPA	48
Db	280	GPPGPAEEGKRARGEGFTGLFPGERGGSGRFPADGVAGPKGAGERSFGPA	339
Qy	49	GPKSGPEAGRPGEAGLPKAGLTGSPGPDGKTGPPGAGDGRPPGPPGARGQA	108
Db	340	GPKSGPEAGRPGEAGLPKAGLTGSPGPDGKTGPPGAGDGRPPGPPGARGQA	399
Qy	109	GVMGFPQPKGAAGPPGKAGRGVPPGAVGPAKDCGAGAGPPGAGPAGERGEQGA	168
Db	400	GVMGFPQPKGAAGPPGKAGRGVPPGAVGPAKDCGAGAGPPGAGPAGERGEQGA	459
Qy	169	GSPGFGGLPGPAGPPGAGKPGGQGVPGDLGAPGSGPAGE-----PGP----	212
Db	460	GSPGFGGLPGPAGPPGAGKPGGQGVPGDLGAPGSGPAGE-----PGP----	519
Qy	213	-----TGLPFGPPGSGRSGRFPADGVAGPKGAGERSFGPAGPR	519
Db	520	GANGAPNDGAKGDAGAPGSGRSGRFPADGVAGPKGAGERSFGPAGPR	579
Qy	259	GE-----AGRPGEAGLPKAGLTGSPGPDGKTGPPGAGDGRP	300
Db	580	GKDGVRGLTGPIGPPGAGAPGSGRSGRFPADGVAGPKGAGERSFGPAGPR	639
Qy	301	GPPGPPGARGQ-----AGVMGFPKGAAGEPKAGRGVPPGPA-----	341
Db	640	GADGQFGAKGEPDAGAKGDAGPPGAGPPGPIGNVGAFCAGKAGSAGPPGATGP	699
Qy	342	-----VGPAGKDCGAGAGPPGAGP-----AGRGEGQAGSPGFGLPAGPPGAGKP	393
Db	700	GAAGRVGPPGPPSGNAGPPGPPGAGKGGKPRGETGPAAGRGVPPGPPGAGKSP	759
Qy	394	GEQGVPGDLGAPGSPGAPGPTGLPFPGERGGSGRFPADGVAGPKGAGERSG	453
Db	760	GADGPAAGPTPGFQGIAGQGVVGLPQGRGSGRFPGLFPGSGEPKQGPSGASGERGP	819
Qy	454	GPAGP---KGSPEAGRPGEAGLPKAGLTGSPGPDGKTGPPGAGDGRPPGPP	510
Db	820	GPWGPPLAGPPGSGREGAPGAGSGRDRGFGAKGDRGETGPAAGPPGAPGAPGV	879
Qy	511	GARGQGVVGMFPKGAAGEPKAGRGVPPGAVGPAKDCG-----AGAQ	558
Db	880	GPAGKSGDRGETGPAAGPVGPAAGRGVPPGQPRGDKGETGEQDGRGKGRGFSGLQ	939
Qy	559	GPPGP-----AGPAGERGEQAGSP-----GFGGLPAGPPGEGAKGPGQVP	603
Db	940	GPPGPPGSGEGQPSGASGAPGPRPGSAGAPGKDLGLPGLPTGPPGPRGTGADGV	999
Qy	604	GDLGAPGSGGAP	616
Db	1000	GPPGPPGPPGPG	1012

Search completed: June 17, 2005, 15:35:34
 Job time : 72.5856 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:42:55 ; Search time 21.1205 Seconds
(without alignments)
2810.812 Million cell updates/sec

Title: US-10-658-989A-2

Perfect score: 3488

Sequence: 1 GPPGEPGPTGLPFPGERGG.....GEQGVPGDLGAPGSPGAGG 617

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2228.5	63.9	1464	1 CGHUIS	collagen alpha 1(I)
2	2205.5	63.2	1453	2 S21626	collagen alpha 1(I)
3	2151.5	61.7	1042	1 CGHUIS	collagen alpha 1(I)
4	2011.5	57.7	671	1 CGRT1S	collagen alpha 1(I)
5	1944	55.7	1418	2 T45467	collagen alpha 1(I)
6	1943	55.7	1487	1 CGHU6C	collagen alpha 1(I)
7	1932	55.4	1419	2 A41182	collagen alpha 1(I)
8	1932	55.4	1487	2 B41182	collagen alpha 1(I)
9	1898	54.4	779	1 CGB01S	collagen alpha 1(I)
10	1890.5	54.2	1486	1 B40333	collagen alpha 1(I)
11	1883	54.0	1492	2 A40333	collagen alpha 1(I)
12	1857.5	53.3	1464	2 S59856	collagen alpha 1(I)
13	1856	53.2	1049	1 CGB07S	collagen alpha 1(I)
14	1853	53.1	1466	1 CGHU7L	collagen alpha 1(I)
15	1840	52.8	886	2 I50694	collagen alpha 1(I)
16	1822.5	52.3	1496	1 CGHU2V	collagen alpha 2(V)
17	1804	51.7	1497	2 I49607	procollagen type V
18	1788.5	51.3	1373	1 A43291	collagen alpha 2(I)
19	1764	50.6	1366	1 CGHU2S	collagen alpha 2(I)
20	1702.5	48.8	1838	1 CGHU1V	collagen alpha 1(V)
21	1689.5	48.4	1843	2 S18803	collagen alpha 1(V)
22	1644	47.1	1806	1 CGHUIE	collagen alpha 1(X)
23	1640.5	47.0	1414	1 S23809	collagen alpha 2(I)
24	1612	46.2	1027	2 S28774	collagen alpha 2(I)
25	1587	45.5	1024	2 S18251	collagen alpha 1(X)
26	1585	45.4	1546	1 CGHU2E	collagen alpha 2(X)
27	1584	45.4	1691	1 S22917	collagen alpha 5(I)
28	1583.5	45.4	673	1 CGB06C	collagen alpha 1(I)
29	1568	45.0	2944	2 A54849	collagen alpha 1(V)

30 1556.5 44.5 1549 2 I48103
31 1550.5 44.5 888 2 S28791
32 1547 44.4 1690 1 CGHUIB
33 1530 43.9 1763 2 S16366
34 1526 43.8 964 1 CGCH2S
35 1522 43.6 1669 1 CGHU4B
36 1510 43.3 920 2 A45748
37 1508 43.2 1603 2 S23810
38 1503.5 43.1 1669 1 CGMS4B
39 1489.5 42.7 1670 1 CGHU3B
40 1481 42.5 1712 1 CGHU2B
41 1471.5 42.2 812 2 S31521
42 1463 41.9 1759 2 T29351
43 1459.5 41.8 1758 2 T29350
44 1456 41.7 1142 2 JX0369
45 1437.5 41.2 1752 2 A45407

ALIGNMENTS

RESULT 1

CGHUIS

collagen alpha 1(I) chain precursor - human

N:Alternate names: procollagen alpha 1(I) chain

C:Species: Homo sapiens (man)

C:Date: 12-Aug-1981 #sequence_revision 04-Oct-1996 #text_change 09-Jul-2004

C:Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S11:

5269; A29439; I53466; A02852; I37247

R:D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.

Gene 67, 105-115, 1988

A:Title: Complete nucleotide sequence of the region encompassing the first twenty-five ex

A:Reference number: I60114; MUID:88329734; PMID:2843432

A:Accession: I60114

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-369, 'L', 371-589 <DAL>

A:Cross-references: UNIPROT:P02452; UNIPROT:Q14992; UNIPROT:Q16053; UNIPROT:Q13896; UNIP

R:Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Procke

Biochem. J. 253, 919-922, 1988

A:Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human t

A:Reference number: S01143; MUID:89025644; PMID:3178743

A:Accession: S01143

A:Molecule type: mRNA

A:Residues: 1-472 <TRO>

A:Cross-references: EMBL:X07884; NID:G30015; PIDN:CAA30731.1; PID:G30016; GB:M36546; NID:

A:Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988

R:Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.;

Nature 310, 337-340, 1984

A:Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of

A:Reference number: A93335; MUID:84270697; PMID:6462220

A:Accession: A93335

A:Molecule type: DNA

A:Residues: 1-58, 'Q', 60-181 <CHU>

A:Cross-references: EMBL:X00820; NID:G35857; PIDN:CAA25394.1; PID:G35658

R:Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.;

J. Biol. Chem. 263, 15151-15157, 1987

A:Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enha

A:Reference number: I55254; MUID:88033098; PMID:2822714

A:Accession: I55254

A:Status: translation not shown; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-45 <ROS>

A:Cross-references: GB:J02829; NID:G180387; PIDN:AAA51993.1; PID:G180388

R:Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gellinas, R.E.

Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987

A:Title: Regulatory elements in the first intron contribute to transcriptional control of

A:Reference number: A39943; MUID:88097389; PMID:3480516

A:Accession: A39943

A:Molecule type: DNA

A:Residues: 1-34 <BOR>

A:Cross-references: GB:J03559; NID:G180876; PIDN:AAA52052.1; PID:G553238

R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

A:Status: translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1187-1194, 'C', 1196-1220 <CH>

A:Cross-references: GB:M3213; NID:G340842; PIDN:AA59363.1; PID:G499622

A:Note: mutant sequence from a patient with mild osteogenesis imperfecta

R:Maekelae, J.K.; Raassina, M.; Virta, A.; Vuorio, E.

Nucleic Acids Res. 16, 349, 1988

A:Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

Query Match 63.9%; Score 2228.5; DB 1; Length 1464;

Best Local Similarity 60.0%; Pred. No. 4.2e-105;

Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

```
QY 1 GPP-----GEPGTCGLPGRGPGSRGPGGADGVAGPKPGAGRGSPGPA 48
   |||
Db 458 GPPGPAGBEGKRGARGEPTGLPGRGPGSRGPGGADGVAGPKPGAGRGSPGPA 517
   |||
QY 49 GPKSGPGBAGPGEAGLPCAKGLTSGSPGPDGKTPPGAGQDGRGPPGPPCARQQA 108
   |||
Db 518 GPKSGPGBAGPGEAGLPCAKGLTSGSPGPDGKTPPGAGQDGRGPPGPPCARQQA 577
   |||
QY 109 GVMGPPGPKGAAGEPKAGRGVPPGAVGAGKDGAGAGQPPGAPGAGEGQGPA 168
   |||
Db 578 GVMGPPGPKGAAGEPKAGRGVPPGAVGAGKDGAGAGQPPGAPGAGEGQGPA 637
   |||
QY 169 GSPGQGLPGAPGPGGAGKGEQVPGDLGAPGSPGAGE-----PGP----- 212
   |||
Db 638 GSPGQGLPGAPGPGGAGKGEQVPGDLGAPGSPGARGRGPPGPPGPPGAPGR 697
   |||
QY 213 -----TGLPQPPGERGPGSRGPGGADGVAGPKPGAGRGSPGCPAGPKGSP 258
   |||
Db 698 GAGAPGNDGAKGDAGAPGAGQAGLQGMGPGERGAAGLPKPGKDGADGPKGADGSP 757
   |||
QY 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGPPGDKTGPFGAGQDGRP 300
   |||
Db 758 GKGVGLTGTPIGPPGAGADKESGSPGAGTGAAGAPGDRGEPGPPGAGFAGPP 817
   |||
QY 301 GPPGPPGARGQ-----AGVMGPPGPKGAAGBPKAGRGVPGPPGA----- 341
   |||
Db 818 GADGQFGAKGEPGDAGAKGDAGPPGAPGPPGPIGNVAGPAGKARGSPGATGFP 877
   |||
QY 342 -----VGPAGKDGAGAGQPPGAPG-----AGERGQAGSPGFGGLPGAPGGEAGKP 393
   |||
Db 878 GAAGRVGPPGSGNAGPPGPPGAPGKGGKGRGTGAGRPGEVPPGPPGAGEKSP 937
   |||
QY 394 GEGVFGDLGAPGSPGAGEPGTGLPGRGPGSRGPGGADGVAGPKPGAGRGSP 453
   |||
Db 938 GADGPAGAGTGPQGIAGRGVGLPGORGERFPGLPGSPGPKQSPGASGERGPP 997
   |||
QY 454 GPAGP---KGSPGEGAGRGAGLPGAKGLTSGSPGPDGKTPPGAGQDGRGPPGPP 510
   |||
Db 998 GPMGPPGLAGPPGSGREGAPGAEKSGRDSGPAKGDGRGTGAPGPPGAXGAGAPGV 1057
   |||
QY 511 GARGQAGVMPGPPGKAGGECAGRGVPPGAVGAGKDG-----AGAQ 558
   |||
Db 1058 GPAGKSGDRGTGAPGAPGVGPPGAPGPPGQGRGDKGTGEGDGRGKRGHGFSLQ 1117
   |||
QY 559 GPPGP-----AGPAGERGEGQGPAGSP---GFQGLPGAPGPGGAGKPGGEGQVP 603
   |||
Db 1118 GPPGPPSGEGQPGSAGSPAGPPGPGSAGAPKDXGLNLGPGIIPGPPGRGTGDPV 1177
   |||
QY 604 GDLGAPGSPGAP 616
   |||
Db 1178 GPPGPPGPPGPPG 1190
   |||
```

RESULT 2

S21626

collagen alpha 1(I) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text change 09-Jul-2004

C:Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626

R:Li, S.W.; Khillan, J.; Prockop, D.J.

Matrix Biol. 14, 593-595, 1994

A:Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I

A:Reference number: S57243

A:Accession: S57243

A:Molecule type: mRNA

A:Residues: 1-1453 <LIS>

A:Cross-references: UNIPROT:P11087; EMBL:U08020; NID:G470673; PIDN:AAA8912.1; PID:G470673

R:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

Biochim. Biophys. Acta 1089, 241-243, 1991

A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.

A:Reference number: S16176; MUID:91274355; PMID:2054384

A:Accession: S16374

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1442-1453 <MET>

A:Cross-references: EMBL:X57981; NID:G50484; PIDN:CAA41046.1; PID:G50485

R:French, B.T.; Lee, W.H.; Maul, G.G.

Gene 39, 311-312, 1985

A:Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.

A:Reference number: A23982; MUID:86137403; PMID:3841523

A:Accession: A23982

A:Molecule type: mRNA

A:Residues: 518-1128 <FR>

A:Cross-references: GB:M14423; NID:G192261; PIDN:AAA37333.1; PID:G192262

R:Monson, J.M.; Friedman, J.; McCarthy, B.J.

Mol. Cell. Biol. 2, 1362-1371, 1982

A:Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for

A:Reference number: I49559; MUID:83141374; PMID:6298597

A:Accession: I49559

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 735-1130 <RR>

A:Cross-references: GB:M17491; NID:G192263; PIDN:AAA37334.1; PID:G192264

R:Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.

Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984

A:Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads to

A:Reference number: I49557; MUID:84170331; PMID:6324198

A:Accession: I49557

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-25 <RE>

A:Cross-references: GB:K01688; NID:G192246; PIDN:AAA37330.1; PID:G553881

R:Penton, S.P.; Lamande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.P.

Biochim. Biophys. Acta 1216, 469-474, 1993

A:Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.

A:Reference number: S39789; MUID:94092741; PMID:8268229

A:Accession: S39789

A:Molecule type: DNA

A:Residues: 1-80, 'E', 82-105, 'D', 107-185, 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-13

R:Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.

Mol. Cell. Biol. 14, 5950-5960, 1994

A:Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indirect

A:Reference number: I48300; MUID:94344105; PMID:8065328

A:Accession: I48300

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>

A:Cross-references: EMBL:X54876; NID:G50486; PIDN:CAA38657.1; PID:G50487

C:Genetics:

A:Gene: COL1A1

A:introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/3

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C:Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple helix;

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-151/Domain: amino-terminal propeptide #status predicted <PRO>

F:30-89/Domain: von Willebrand factor type C repeat homology <WVC>

F:152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>

F:1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match

Best Local Similarity 63.2%; Score 2205.5; DB 2; Length 1453;

Matches 432; Conservative 22; Mismatches 162; Indels 117; Gaps 11;

QY 1 GPP-----GEPGPTGLPGPPGGRGGSGRGGPGADGVAGPKGPPAGERGSPGA 48
 Db |||||
 447 GPPGAGBEGKRGARCEGPGSLPDPGGRGGSGRGGPGADGVAGPKGPPAGERGSPGA 506
 QY 49 GPKGSPGAGRGEAGLPGAKGLTGSPPGDPDKTTPPGAGQDGRPPGPPGARGQA 108
 Db |||||
 507 GPKGSPGAGRGEAGLPGAKGLTGSPPGDPDKTTPPGAGQDGRPPGPPGARGQA 566
 QY 109 GVMGPPGKGAAGEPGKAGRGVPPGAVGPKAGDKGAGAGQGGPPGAGPAGERGEGQA 168
 Db |||||
 567 GVMGPPGKGTAGEPGKAGRGVPPGAVGPKAGDKGAGAGQGGPPGAGPAGERGEGQA 626
 QY 169 GSPGQGLPGPAGPGEAGKGEQGVQPDILGAPGSPGAGE-----PGP----- 212
 Db |||||
 627 GSPGQGLPGPAGPGEAGKGEQGVQPDILGAPGSPGAGE-----PGP----- 686
 QY 213 -----TGLPGPGERGGSGRGGPGADGVAGPKGPPAGERGSPGAGPR 258
 Db |||||
 687 GNNGAPGNDGAKGDTGAFCAFGSQAGPQLQMPGERRGAAGLPGPKGDRGADGPKGADGSP 746
 QY 259 GE-----AGRPGEAGLPGAKGLTGSPPGPD-----GKTGPPGAGQDGRP 300
 Db |||||
 747 GKDGAAGLGTGIPGPPGAGAPGDKGAGSGPPGPTGARGAPGDRGEGAGPPGAGP 806
 QY 301 GPPGPPGARGO-----AGVMGFPDPCGAAGEPKAGRGVPPGPA----- 341
 Db |||||
 807 GADQPGAKGPPGDTGVKGDAAGPPGPPGAGPPGPIGNVAGPFGKPRGAAGPPGATGFP 866
 QY 342 -----VGPAGKDEBAGAQGPPGAGP-----AGERGEQGPAGSPGQGLPGPAGPPGAGKP 993
 Db |||||
 867 GAAGRVGPPGSGNAGPPGPPGVGKGGKPRGCTGAPRGVEVPPGPPGAGKGGSP 926
 QY 394 GEQGVQDILGAPGSPGAGPPTGLPDPGGRGGSGRGGPGADGVAGPKGPPAGERGSP 453
 Db |||||
 927 GADGAPGSPGTPGQGTAGQGVVGLPGQGRGGRFPGLPGSPGEPKQGFSGSSGGRGPP 986
 QY 454 GPAGP---KGSPEAGRPGAGLPGAKGLTGSPPGDPDKTTPPGAGQDGRPPGPPG 510
 Db |||||
 987 GPMGPPGLAGPPGESREGSPGAGSFGDRGAPCAKAGDRGCTGAPGPPGAPGAPGVP 1046
 QY 511 GARGQAGVMGFPDPKGAAGBPGKAGRGVPP-----PGAV 546
 Db |||||
 1047 GPAGKNGDRGCTGAPGAPGIPGAGARGPAGPQGRGDKGTGEGQDRIKGRHFGSLQ 1106
 QY 547 GPAGKDEAGAGQGPAGPAGRGGRGQAGSP---GFOGLPGPAGPPGAGKPOEQGV 603
 Db |||||
 1107 GPPGSPGSPGQGPAGSAGPAGPPGPPGAGSPGKDGGLNGLPGLPPIGPPGPRGRTGDSGA 1166
 QY 604 GDLGAPGSPGAG 616
 Db |||||
 1167 GPPGPPGPPGPPG 1179

RESULT 3
 CGCH15
 collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
 C;Species: Gallus gallus (chicken)
 C;Date: 12-Aug-1981 #sequence revision 06-Jul-1982 #text_change 31-Mar-2000
 C;Accession: A90458; A90181; A02857
 R;Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.
 Biochemistry 21, 2048-2055, 1982
 A;Title: Amino acid sequence of chick skin collagen alpha1(I)-CB8 and the complete prime
 A;Reference number: A90458; MUID:82231995; PMID:7093229
 A;Accession: A90458
 A;Molecule type: Protein
 A;Residues: 1-1036 <HIG>
 A;Experimental source: skin
 A;Note: this is the latest in a series of papers from these workers elucidating the sequ
 R;Eyre, D.R.; Glimcher, M.J.
 Biochem. Biophys. Res. Commun. 48, 720-726, 1972
 A;Title: Evidence for a previously undetected sequence at the carboxyterminus of the alp
 A;Reference number: A90181; MUID:72243016; PMID:5047697
 A;Accession: A90181

A;Molecule type: protein
 A;Residues: 1037-1042 <EYR>
 A;Experimental source: skin
 A;Note: residues 1037-1042 above correspond to the carboxyl end of the protein

C;Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some
 C;Comment: Most of the prolines at the third position of the tripeptide repeating unit (C
 C;Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in pos
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyrrolidone carboxylic acid; trimer; t
 F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 61.7%; Score 2151.5; DB 1; Length 1042;
 Best Local Similarity 57.8%; Pred. No. 2.4e-101;
 Matches 424; Conservative 17; Mismatches 175; Indels 117; Gaps 10;

QY 1 GPP-----GEPGPTGLPGPPGGRGGSGRGGPGADGVAGPKGPPAGERGSPGA 48
 Db |||||
 296 GPPGAGBEGKRGARCEGPGSLPDPGGRGGSGRGGPGADGVAGPKGPPAGERGSPGA 355
 QY 49 GPKGSPGAGRGEAGLPGAKGLTGSPPGDPDKTTPPGAGQDGRPPGPPGARGQA 108
 Db |||||
 356 GPKGSPGAGRGEAGLPGAKGLTGSPPGDPDKTTPPGAGQDGRPPGPPGARGQA 415
 QY 109 GVMGPPGKGAAGEPGKAGRGVPPGAVGPKAGDKGAGAGQGGPPGAGPAGERGEGQA 168
 Db |||||
 416 GVMGPPGKGAAGEPGKAGRGVPPGAVGPKAGDKGAGAGQGGPPGAGPAGERGEGQA 475
 QY 169 GSPGQGLPGPAGPGEAGKGEQGVQPDILGAPGSPGAGE-----PGP----- 212
 Db |||||
 476 GSPGQGLPGPAGPGEAGKGEQGVQPDILGAPGSPGAGE-----PGP----- 535
 QY 213 -----TGLPGPGERGGSGRGGPGADGVAGPKGPPAGERGSPGAGPR 251
 Db |||||
 536 GADQPGAKGPPGDTGVKGDAAGPPGPPGAGPPGPIGNVAGPFGKPRGAAGPPGATGFP 595
 QY 252 -----AGPKSPGABRPP---GEAGLPGAKGLTGSPPGDPDKTTPPGAGP----- 294
 Db |||||
 596 GKDGLRGLTGPFGPPGAGAPGDKGAGPPGAGTGAARGAPGDRGEPGPPGAGPAGPP 655
 QY 295 -----GQDGRPPGPPGARGQAGVMGFPDPKGAAGBPGKAGRGVPP 336
 Db |||||
 656 GADQPGAKGPPGDTGVKGDAAGPPGPPGAGPPGPIGNVAGPFGKPRGAAGPPGATGFP 715
 QY 337 GPPGAVGPKAGDKGAGAGQGPAGPAGPAGP---RGEQGPAGSPGQGLPGPAGPPGAGKP 393
 Db |||||
 716 GAAGRVGPPGSGNIGLPGPPGAGKZGSKGRGCTGAPRGVEVPPGPPGAGKGGSP 775
 QY 394 GEQGVQDILGAPGSPGAGPPTGLPDPGGRGGSGRGGPGADGVAGPKGPPAGERGSP 453
 Db |||||
 776 GADGPIGAPGTPGQGTAGQGVVGLPGQGRGGRFPGLPGSPGEPKQGFSGSGASGERGP 835
 QY 454 GPAGP---KGSPEAGRPGAGLPGAKGLTGSPPGDPDKTTPPGAGQDGRPPGPPG 510
 Db |||||
 836 GPMGPPGLAGPPGESREGSPGAGSFGDRGAPGCAKAGDRGCTGAPGPPGAPGAPGVP 895
 QY 511 GARGQAGVMGFPDPKGAAGBPGKAGRGVPPGPGAVGPPGAGKPOEQGV 558
 Db |||||
 896 GPAGKNGDRGCTGAPGAPGIPGAGARGPAGPQGRGDKGTGEGQDRIKGRHFGSLQ 955
 QY 559 GPP-----GPAGPAGERGEGGAGSPGFGGLPGPAGPPGAGKPPGQGV 603
 Db |||||
 956 GPPGPPGAGGEGGAGSPGAPGPPGPPGAGAGKDGGLNGLPGLPPIGPPGPRGRTGDSGV 1015
 QY 604 GDLGAPGSPGAG 616
 Db |||||
 1016 GPPGPPGPPGPPG 1028

RESULT 4
 CGRI15
 collagen alpha 1(I) chain - rat (tentative sequence) (fragments)
 C;Species: Rattus norvegicus (Norway rat)
 C;Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000

A:Residues: 419-567 <BU3>
A:Experimental source: skin
R:Stoltz, M.; Timpl, R.; Furthmayr, H.; Kuehn, K.
Eur. J. Biochem. 37, 287-294, 1973
A:Title: Structural and immunogenic properties of a major antigenic determinant in neutralized and subsequently O-glycosylated.
A:Reference number: A91209; MUID:74011954; PMID:4126850
A:Contents: CNBr6
A:Accession: A91209
A:Molecule type: protein
A:Residues: 568-651 <ST1>
A:Experimental source: skin
A:Note: this region probably corresponds to positions 949-1032 of the alpha 1(I) chain
A:Note: the major antigenic determinant (of neutral salt-extracted rat skin collagen) in neutralized and subsequently O-glycosylated.
R:Stoltz, M.; Timpl, R.; Kuehn, K.
FEBS Lett. 26, 61-65, 1972
A:Title: Non-helical regions in rat collagen alpha1-chain.
A:Reference number: A91385; MUID:73049495; PMID:4636751
A:Contents: CNBr6
A:Accession: A91385
A:Molecule type: protein
A:Residues: 651-671 <ST2>
A:Experimental source: skin
A:Note: the composition of peptides comprising residues 1-9 and 1-19 confirms the sequence
A:Note: this region (residues 651-671 above) probably corresponds to positions 1032-1052
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C
ed and subsequently O-glycosylated.
C:Comment: The order of the nine CNBr peptides in the alpha 1(I) chain of rat skin collagen
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxyl
F:1/Modified site: blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #status
F:9/Modified site: allysine (Lys) #status experimental
F:103,424,547/Binding site: carbohydrate (Lys) (covalent) #status experimental
F:103/Modified site: 5-hydroxylysine (Lys) #status experimental
F:424,547/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Query Match 57.7%; Score 2011.5; DB 1; Length 671;
Best Local Similarity 59.5%; Pred. No. 1.8e-94;
Matches 396; Conservative 28; Mismatches 119; Indels 123; Gaps 14;

Qy	1	GPPGEPGPTGLPPGGERGGGS-----RGPPGA-----DGVAGPKPACERSGSPGA 48
Db	29	GPPGAPGQGGPQGGPEGPEGPGASGPMGRPGPPGPKNGDDGEAGKPGRPGQGGPPGQ 88
Qy	49	GPKGSPGEAGRPGE-----AGLPCAKGLT-----GSPGSPGD---GKTGPPGPGAGQD 93
Db	89	GARGLPCTAGLPGMKHGRGFSGLDCAKNGTGPAGPKGPGSPGNGAPQCMGRGLPGER 148
Qy	94	GRPGPPGPGARGQAGVWGMFGPGKAAAE-----PGKAGERGVGPPGAVGPAKDGEEA 147
Db	149	GRPGPPGSAGARGDDGAVGAAGPPTGTPGPPGFGAAGKAGEAGPQGARGSEGPQGV 208
Qy	148	GAQPPGPGPAGPAGERGEGP-----AGSPGFOGLPCGAPPPPGAEAGKEGQ 192
Db	209	GEPPGPPGAGAAGPAGNPGADGQPGAKGANGAPGTAGAPFPFGARPGSPGQGSAGPK 268
Qy	193	GVPGDLGAP-----GFSGPA-----GEPGPTGLPAPPGERGGP 225
Db	269	GNSGEPGAPGNKGTGAKGEPGAGVQPPGPAEGEKGARGEPGSLGPPGGERGGP 328
Qy	226	GSRGFPAGDVAGPKPAGERGSGPGAPKSGPSGBAGRPGEAGLPGAKGLTSGSPGSPGD 285
Db	329	GSRGFPAGDVAGPKPAGERGSGPGAPKSGPSGBAGRPGEAGLPGAKGLTSGSPGSPGD 388
Qy	286	GKTGPPGAGDGRPPGPGARGQAGVMGFPFGKGAAGERGPKAGERGVGPPGAVGPA 345
Db	389	GKTGPPGAGZBGRPPGAPPGARGQAGVMGFPFGKGTAGEPGKAGERGVGPPGAVGPA 448
Qy	346	GKDGEAGAQGPPGAPGAGERGEGPAGSPGFGQLPGPAGPPGAEAGKPGEGQVPGDLGAP 405
Db	449	GKDEAGAQGAPGAPGAPGAGERGEGQAGSPGFGQLPGPAGPPGAEAGKPGZVZGVPDLGAP 508
Qy	406	GPSGPAGEPGPTGLPFPGERGGPSRGGFPFGADGVAGPKPAGERGSGPAGPKGSPGEEA 465

Db 509 GFSGARGE-----RGFPGRGVQGPFGPAGPRGN-----NGAPGBB 544
 QY 466 GRPGEAGLPGAKGLTSGPSGPGDGTGTPGPGAGQDGRPGPPGPPGAGQAGVWGFPGPK 525
 Db 545 GAKGDTGAPGAPGSGQAGLZGMSGLZGFPFPGS-----PGZGFSGAGSPAGPR---GPP 598
 QY 526 GAAGRPKAGRGVGPAGVAGPAGKDGAGAGQAGPAGPAGPAGRGEGQAGPAGSPGFQGLP 585
 Db 599 GSAGSPGKBLGLFCPIGPPGPRGRTGBAGSPGSPGPPG-----p 640
 QY 586 GPAGPP 591
 Db 641 GPPGP 646

RESULT 5
 T45467
 collagen alpha 1(II) chain precursor [imported] - horse
 N:Alternate names: type II collagen
 C:Species: Equus caballus (domestic horse)
 C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
 R:Accession: T45467
 R:Richardson, D.W.; Dodge, G.R.
 submitted to the EMBL Data Library, June 1996
 A:Description: Cloning of equine type II collagen and modulation of its expression in eq
 A:Reference number: Z22977
 A:Accession: T45467
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1418 <RIC>
 A:Cross-references: UNIPROT:Q28396; EMBL:U62528; PIDN:AA805773.1
 C:Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 55.7%; Score 1944; DB 2; Length 1418;
 Best Local Similarity 52.1%; Pred. No. 7.3e-91;
 Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;
 QY 1 GPPCEPPTGLGPPGPPGSGRPGGADGAGKPGAGRGSPGAPGKSPGAPGAGRP 60
 Db 156 GNPGEFGEVSGPMGPRGPPGPPGKPGDDGEAGKPKGSGRPPGPPGQARGPFGTGLP 215
 QY 61 -----GBAGLPGAKGLTSP---GSPGP-----DGKTGPPGPA--- 90
 Db 216 GVKGHRGYPLDGAKEAGAPGVKGSFSGENGSPGPPGRLPGERRTGPAAGAR 275
 QY 91 GQDRPDPGPPGARGOAGVWGPFGKGAAGEPCKAGER-----GVPGP- 134
 Db 276 GNDGQPGAPGPPGVGPPAGGPPGAPGAGKGEAGPTGARGPGEAGPGEPTGPGSPGA 335
 QY 135 -----PGAVPAGKDGAGAGQPPGPPGAPGAGERGEGP-----AGSPGFQGLP 177
 Db 336 GAAGNPCTDIPGAKGSAGAPGIAGAPGPPGPPGPGATGPIPGKQGTGEPGIAGFK 395
 QY 178 GPAGPPEAGKPEOGVGRDLGAPSGPAGPPTGTLDPGPRGPGSGRGGFPAGDGA 237
 Db 396 GEGPKEGEPGAPQAPGAPGAEGRGARGPFGAGVPGPPGERGAPGNRGGFPQDGLA 455
 QY 238 GPKPAGERGSPGAPGKSPGEPGAGLPAGKGLTSGSPGPPGDKTGPFGPAGQD 297
 Db 456 GPKGAPGERGSLAGPKANGDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 515
 QY 298 GRPDPGPPGARGOAGVWGFPGKGAAGPAGRGVPPGPPGAVGPPGKDGAGAGAGQPP 357
 Db 516 GRPDPGPPGARGOAGVWGFPGKGAAGPAGRGVPPGPPGPPGPPGPPGPPGPPGPPG 575
 QY 358 GPAGPAGERGEGPAGSPGFQGLPAGPAGPAGKPEOGVGRDLGAPGSPGAGE---p 414
 Db 576 GPAGPAGERGEGPAGSPGFQGLPAGPAGPAGKPEOGVGRDLGAPGSPGAGE---p 635
 QY 415 GTGLP-----GPPGERGPGSGRGGPFGDAGVAGPKGA 447
 Db 636 GERSGFGAQLQCARGLPCTGTGPKGASGAPGAPGAGPPGAGPPGLQGWGEGRGAAGIAGPK 695

QY 448 GERGSPGAPGPKSGPGE-----AGRPGEAGLPGAKGLTSGPSGPGPD 489
 Db 696 GDRGDVGEKGPGEAGKDGKGRGLTGPIGPPGAGANGKEGVEGPPGAGTACAGAPGR 755
 QY 490 GKTGPPGAPAGQDRPCPCPPGARG-----GQAVWGPFGPKGAAGEP-----GKA 534
 Db 756 GTGTPPPGAPGAPGADGQPGKAGEQGEAGQKDGAGAPGPPGSGARGPQGTGTGTPK 815
 QY 535 GERGVPPGPGA-----VGPAGKDGAGAGQAGPPGPA---GPAGERGEGQAGSPGPG 582
 Db 816 GARGAGQPPGATGFPGAAGRVGPPGSGNGNPPGPPGPPGSGKDGKAGKAGDSGPPGRAGDP 875
 QY 583 GLPGPAGPPGEGAGKPEOGVPPGDLGAPGSPGAG 616
 Db 876 GLQGPAGPPGEGKGEFGDDGSPGDPGPPGPPGQGLAG 909

RESULT 6

CGHU6C

collagen alpha 1(II) chain precursor [validated] - human
 N:Alternate names: procollagen alpha 1(II) chain
 N:Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collagen e
 C:Species: Homo sapiens (man)
 C:Date: 28-May-1986 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
 C:Accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S63;
 7250; I37251; I37252; I37253; I37254; I53338; I59535; I61910
 R:Ryan, M.C.; Sieraski, M.; Sandell, L.J.
 Genomics 8, 41-48, 1990
 A:Title: The human type II procollagen gene: identification of an additional protein-codi
 A:Reference number: A38513; MUID:91184811; PMID:2081599
 A:Accession: A38513
 A:Molecule type: DNA
 A:Residues: 1-103 <RYA>
 A:Cross-references: UNIPROT:P02458; UNIPROT:Q14042; UNIPROT:Q16672; UNIPROT:Q12985; UNIP
 R:Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
 Nucleic Acids Res. 17, 9473, 1989
 A:Title: Nucleotide sequence of the full length cDNA encoding for human type II procolla
 A:Reference number: S06715; MUID:90067946; PMID:2587267
 A:Accession: S06715
 A:Molecule type: mRNA
 A:Residues: 1-28, 'R', 99-1487 <SU2>
 A:Cross-references: EMBL:X16468; NID:g29515; PIDN:CAA344488.1; PID:g29516
 A>Note: alternative splice form 1
 R:Vikkula, M.; Metsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.
 Biochem. J. 285, 287-294, 1992
 A:Title: Structural analysis of the regulatory elements of the type-II procollagen gene.
 A:Reference number: S24270; MUID:92344585; PMID:1637314
 A:Accession: S24270
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-28 <VIK>
 A:Cross-references: EMBL:X58709; GB:S40537; NID:g35659
 A>Note: this translation is not annotated in GenBank entry HSPROCOE1, release 111.0
 R:Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
 Gene 44, 11-16, 1986
 A:Title: Promoter region of the human pro-alpha-1-(II)-collagen gene.
 A:Reference number: A24828; MUID:87031574; PMID:3021582
 A:Accession: A24828
 A:Molecule type: DNA
 A:Residues: 1-8, 'T', 10-28 <NUN>
 A:Cross-references: GB:M25698; NID:g180872; PIDN:AAA52051.1; PID:g553237
 R:Baldwin, C.T.; Reginato, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
 Biochem. J. 262, 521-528, 1989
 A:Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(II) c
 A:Reference number: S06496; MUID:90026318; PMID:2803268
 A:Accession: S06496
 A:Molecule type: mRNA
 A:Residues: 7-28, 'R', 99-157, 'P', 159-440, 'G', 442-456, 'E', 458-640, 'A', 642-831, 'PA', 834, 'F',
 A:Cross-references: EMBL:X16711; NID:g30040; PIDN:CAA34683.1; PID:g30041
 A>Note: alternative splice form 1
 R:Ryan, M.C.; Sandell, L.J.
 J. Biol. Chem. 265, 10334-10339, 1990

A>Title: Differential expression of a cysteine-rich domain in the amino-terminal propeptide
A:Reference number: A35428; MUID:90285153; PMID:2355003
A:Accession: A35428
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 27-81,'L',83-103 <RYA2>
A>Note: alternative splice form 2; splicing appears to be under developmental regulation
R:Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
Genomics 4, 438-441, 1989
A>Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide conf
A:Reference number: A30147; MUID:89233138; PMID:2714801
A:Accession: A30147
A:Molecule type: DNA
A:Residues: 104-157,'P',159-236 <SUM>
A:Cross-references: GB:J03065; GB:M23660; GB:M25655; GB:M25730; GB:M32168; GB
R:Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990
A>Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pri
A:Reference number: A94227; MUID:90370826; PMID:1975693
A:Accession: A33116
A:Molecule type: DNA
A:Residues: 171-172,'C',174-175 <ALA>
A>Note: mutant sequence from a family with primary generalized osteoarthritis
R:Diab, M.; Wu, J.J.; Eyre, D.R.
Biochem. J. 314, 327-332, 1996
A>Title: Collagen type IX from human cartilage: a structural profile of intermolecular c
A:Reference number: S64673; MUID:96195147; PMID:8660302
A:Accession: S64674
A:Molecule type: protein
A:Residues: 188-189,'X',191-195;1224-1230,'X',1232-1236 <DIA>
R:Franc, S.; Marzin, E.; Bouthillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,
Eur. J. Biochem. 234, 125-131, 1995
A>Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil car
A:Reference number: S63514; MUID:96096730; PMID:8529631
A:Accession: S63514
A:Molecule type: protein
A:Residues: 243-261;575-590;756-763,'X',765-779 <FRA>
R:Tiller, G.E.; Weis, M.A.; Polumbo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre,
Am. J. Hum. Genet. 56, 388-395, 1995
A>Title: An RNA-splicing mutation (G+5IVS20) in the type II collagen gene (COL2A1) in a
A:Reference number: I38867; MUID:95150028; PMID:7847372
A:Accession: I38867
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 440,'G',442-456,'E',458-480,'P',482-509 <TIL1>
A:Cross-references: EMBL:U15195; NID:9557053; PIDN:AAB60370.1; PID:9557054
R:Ramirez, F.
Submitted to the EMBL Data Library, December 1988
A:Reference number: S04892
A:Accession: S04892
A:Molecule type: mRNA
A:Residues: 501-676,'A',678-783,'A',785-831,'PA',834,'P',836-1214 <RAM>
A:Cross-references: EMBL:X13783; NID:G30037; PIDN:CAA32030.1; PID:g930050
R:Vikkula, M.; Peltonen, L.
FEBS Lett. 250, 171-174, 1989
A>Title: Structural analyses of the polymorphic area in type II collagen gene.
A:Reference number: S05000; MUID:89325561; PMID:2753125
A:Accession: S05000
A:Molecule type: DNA
A:Residues: 630-640,'A',642-785 <VIK2>
A:Cross-references: EMBL:X16158; NID:G29951; PIDN:CAA34278.1; PID:g1335018; PIDN:CAA3427
PIDN:CAA34283.1; PID:g1335023; PIDN:CAA34284.1; PID:g1335024
R:Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D
J. Biol. Chem. 267, 22522-22526, 1992
A>Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain pro
A:Reference number: A44309; MUID:93054548; PMID:1425602
A:Accession: A44309
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA; mRNA
A:Residues: 752-831,'PA',834,'P',836-1005,'K',1007-1036,'O',1038-1052,'E',1054-1068,'T',
A:Cross-references: GB:L00977; NID:g180812; PIDN:AAB23314.1; PID:g258774
A>Note: sequence extracted from NCBI backbone (NCBIP:117273); parts of this sequence we
A:Note: this translation is not annotated and this publication is not cited in GenBank e

A>Note: mutant sequence associated with perinatal lethal hypochondrogenesis
R:Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
A>Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual
A:Reference number: S16502; MUID:90251662; PMID:2339128
A:Accession: S16502
A:Molecule type: DNA
A:Residues: 1164-1184,'GPSGKGANGIEGPT',1185-1199 <TIL2>
A:Cross-references: EMBL:M37126; NID:g180808; PIDN:AAA52037.1; PID:g180809
A>Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
R:Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
A>Title: Identification and characterization of the human type II collagen gene (COL2A1)
A:Reference number: A02858; MUID:85190534; PMID:3857598
A:Accession: A02858
A:Molecule type: DNA
A:Residues: 1032-1056,'N',1058-1068,'T',1070-1487 <CHE>
A:Cross-references: GB:J00116; NID:g180395; PIDN:AAA51997.1; PID:g180396
R:Elima, K.; Vuorio, T.; Vuorio, E.
Nucleic Acids Res. 15, 9499-9504, 1987
A>Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) co
A:Reference number: A27280; MUID:88067771; PMID:2825137
A:Accession: A27280
A:Molecule type: DNA; mRNA
A:Residues: 1175-1487 <ELI>
A:Cross-references: EMBL:X06268; NID:g30096; PIDN:CAA29604.1; PID:g30097
A:Experimental source: fetal epiphyseal cartilage
R:van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
Biochem. J. 237, 923-925, 1986
A>Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
A:Reference number: A57033; MUID:87099927; PMID:3800925
A:Accession: A57033
A:Molecule type: protein
A:Residues: 'XE',1244-1246,'N',1248,'X',1250-1265;1295-1305;1395-1408 <VAN>
A>Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal prop
R:Strom, C.M.; Upholt, W.B.
Nucleic Acids Res. 12, 1025-1038, 1984
A>Title: Isolation and characterization of genomic clones corresponding to the human type
A:Reference number: A21733; MUID:84118798; PMID:6320112
A:Accession: A21733
A:Molecule type: DNA
A:Residues: 1245-1295 <STR1>
A:Cross-references: EMBL:X00339; EMBL:X00298; NID:g394699; PIDN:CAA25092.1; PID:g4378975
A:Accession: B21733
A:Molecule type: DNA
A:Residues: 894-909,'PE', <STR2>
A:Cross-references: GB:X01785; NID:g30035; PIDN:CAA25082.1; PID:g1335032
R:Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A>Title: Isolation and partial characterization of genomic clones coding for a human pro
gene.
A:Reference number: A24561; MUID:86104139; PMID:3002437
A:Accession: A24561
A:Molecule type: DNA
A:Residues: 1296-1358 <NUN2>
A:Cross-references: GB:M12048; NID:g180017
A>Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
A>Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with the
R:Sangiorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez,
Nucleic Acids Res. 13, 2207-2225, 1985
A>Title: Isolation and partial characterization of the entire human pro alpha 1(II) colla
A:Reference number: I37249; MUID:85215609; PMID:2987845
A:Accession: S59491
A:Molecule type: DNA
A:Residues: 7-28,'R',99-114;541-578;786-802;1055-1056,'N',1058-1068,'T',1070-1109;1200-11
A:Accession: I84453
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 7-28 <SAN2>
A:Cross-references: GB:M23759; NID:g180845; EMBL:X03320; GB:M24938; NID:g30104
A>Note: the GenBank PID is based on an incorrect reading frame
A:Accession: I37250
A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 541-560 <SAN3>
A:CROSS-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CA26227.1; PID:g929621
A:Accession: I37251

Query Match 55.7%; Score 1943; DB 1; Length 1487;
Best Local Similarity 52.1%; Pred. No. 8.5e-91;
Matches 393; Conservative 34; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGPGTGLPDPGPGGPGSGFPAGADGAGVAGKPGAGGSGPGAGKPGSGFGEAGRP 60
DB 225 GNPGEPPGVSPPGPPGPPGPKPDGDDGAGKPKAGGPPGPGQARGPFGTGLP 284
QY 61 -----GEAGLPGAKGLTGPSP-----GSPGP-----DGKTGPPGPA--- 90
DB 285 GVKHGRGVPGLDAGKAGAGAVGVGSGSPGSGPGMGRGLPGRGRTGAGAGAR 344
QY 91 GQDGRPGPPGPPGARGQAVGMGFP-----GPKAAAGEPKAGRGVPPGAVGPA 141
DB 345 GNDQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 404
QY 142 -----GKDGAGAGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 195
DB 405 GASNPPTDGTGPKAGSAGAPGAGFPFPPGPPGPPGPPGPPGPPGPPGPPGPP 464
QY 196 GD-----LGAPGSPGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 237
DB 465 GEQPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 524
QY 238 GPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 297
DB 525 GPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 584
QY 298 GRPFPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 357
DB 585 GRFPFPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 644
QY 358 GPAGPAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 413
DB 645 GPAGPAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 704
QY 414 -----PGPTGLP-----GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 447
DB 705 GERGSPGAGLQGLPGLPGTGTGDPKAGSPAGPPGAGQPPGLQMPGPPGPPGPP 764
QY 448 GERGS-----PPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 824
DB 765 GDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 884
QY 484 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 534
DB 825 GETGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 884
QY 535 GERGVPPGPPG-----VGPAGKDGAGAGQPPGPPG-----GPKAAAGEPK 582
DB 885 GARGAQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 944
QY 583 GLPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 616
DB 945 GLQPPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 978

RESULT 7
A41182
collagen alpha 1(II) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C:Accession: A41182; A4485
R:MetaeRanta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
A:Reference number: A41182; MUID:91358489; PMID:1885613
A:Accession: A41182

A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1419 <MET>
A:CROSS-references: GB:M65161
R:Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
Development 111, 945-953, 1991
A:Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilage
A:Reference number: A44885; MUID:91347939; PMID:1879363
A:Accession: A44885
A:Molecule type: DNA
A:Residues: 1-28 <CHE>
A:CROSS-references: GB:S63190; NID:g234368; PIDN:AAB19627.1; PID:g234369
A:Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBI:63192)
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
F:1191-1419/Domain: fibrillar collagen carboxyl-terminal homology <PCC>

Query Match 55.4%; Score 1932; DB 2; Length 1419;
Best Local Similarity 51.9%; Pred. No. 2.9e-90;
Matches 391; Conservative 36; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGPGTGLPDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 60
DB 157 GNPGEPPGVSPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 216
QY 61 -----GEAGLPGAKGLTGPSP-----GSPGP-----DGKTGPPGPA--- 90
DB 217 GVKHGRGVPGLDAGKAGAGAVGVGSGSPGSGPPGPPGPPGPPGPPGPPGPPG 276
QY 91 GQDGRPGPPGPPGARGQAVGMGFP-----GPKAAAGEPKAGRGVPPGAVGPA 141
DB 277 GNDQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 336
QY 142 GKDS-----AGAOPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 177
DB 337 GASNPPTDGTGPKAGSAGAPGAGFPFPPGPPGPPGPPGPPGPPGPPGPPGPP 396
QY 178 GPAGPAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 237
DB 397 GDQPKGETGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 456
QY 238 GPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 297
DB 457 GPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 516
QY 298 GRPFPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 357
DB 517 GRFPFPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 576
QY 358 GPAGPAGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 413
DB 577 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 636
QY 414 -----PGPTGLP-----GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 447
DB 637 GERGSPGAGLQGLPGLPGTGTGDPKAGSPAGPPGAGQPPGLQMPGPPGPPG 696
QY 448 GERGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 489
DB 697 GDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 756
QY 490 GKTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 534
DB 757 GETGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 816
QY 535 GERGVPPGPPG-----VGPAGKDGAGAGQPPGPPG-----GPKAAAGEPK 582
DB 817 GARGAQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 876
QY 583 GLPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 616
DB 877 GLQPPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 910

RESULT 8

B41182
C:Species: Mus musculus (house mouse)
C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 09-Jul-2004
C:Accession: B41182
R:Metaerantca, M.; Tonan, D.; de Crombrughe, B.; Vuorio, E.
J. Biol. Chem. 266, 15862-15869, 1991
A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
A:Reference number: A41182; MUID:91358489; PMID:1885613
A:Accession: B41182
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1487 <MET>
A:Cross-references: UNIPROT:Q62031; UNIPROT:Q62032; UNIPROT:Q62033; GB:M65161
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
F:33-91/Domain: von Willebrand factor type C repeat homology <VMC>
F:1259-1487/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 55.4%; Score 1932; DB 2; Length 1487;
Best Local Similarity 51.9%; Pred. No. 3e-90;
Matches 391; Conservative 36; Mismatches 189; Indels 138; Gaps 14;

QY 1 GPPGPGTGLPCPPGCGGSGRFGPCADGVAGPKGAPGAGSGPPGAPGKSGPGEAGRP 60
DB 225 GNPGEPPGVSGPMGPPGPPGAPGKGDGEAGPKSGERGLPGPQGARGFPFTGLP 284

QY 61 -----GEAGLPKAGLGTSP-----GSPGP-----DGKTGPPGPA--- 90
DB 285 GVKHGRVPLDGAKEGAGAPGVKESGPNESGPMGPRGLPGERGTGPAAGAR 344

QY 91 QDGRPGPPGPPGARGQAGVMGFP-----GPKAAGBFGKAGRGVFPGPAGVGA 141
DB 345 GNDGQGPAGPPGVPVAGGPFPCAPCAKEAGTGAAGPGEAGQSGRGEFNGSPGPA 404

QY 142 GKDE-----AGAQPPGAPGAGRGQGP-----AGSPGFQGLP 177
DB 405 GASGNPPTDGI PGAKGSAGAPGIAGAPGPPGPPGQATGPLGPKGQAGEPGIAGF 464

QY 178 GPAGPPGAGKPGEGVPGDLGAPGSPGAPGPTGLPPGPPGSGGSGRFGPGADGVA 237
DB 465 GDQPKGTGTAGPAGAPGAPGAGEGRGARPEGAGPIGPPGKAGPKNRGFPQDGLA 524

QY 238 GPKGAPGERSGPPGAPGKSGPGEAGLPKAGLGTSGSPGPDGKTGPPGPAQD 297
DB 525 GPKGAPGERSGSLAGKAGNDPCRGEPGLPGARGLTGREGDAGPQGVKVPSPGAPGD 584

QY 298 GRPGRPPPGARGQAGVMGFPGPKAAGEPKAGRGVFPGPAGVGPAGKDGGEAGAGQPP 357
DB 585 GRPGRPPPGARGQAGVMGFPGPKGANGEPGKAGKGLAGAPGLGPKDGETGAAGPP 644

QY 358 GPAGPAGRGEGGPPGSGPFGQLGPPAGPGEAGKPGEGVPGDLGAPGSPGAGE----- 413
DB 645 GPSGPAGRGEGGAPGSPGFGQLGPPGPPGEGGQGDGIPGEGAGAPGLVPPRGGERFP 704

QY 414 -----PGPTGLP-----GPPGRRGPGSGRFGPCADGVAGPKGPA 447
DB 705 GERGSFGAGLQGRFLGPTGTTDGPKNAGPDGPPGAGPPGLQGMFGERGAAGIAGPK 764

QY 448 GERGSFGAPGKSGPGE-----AGRPGEAGLPKAGLGTSGSPGSPGPD 489
DB 765 GDRGDVGEKPGEPAGKDGGRGLTGPFGPPGAGANGKEGVEGPPGSPGSGTGARGAPCEP 824

QY 490 GKTGPPGAGQDGRGPPGPPCAR-----GQAGVMGPPGPKGAAGEP-----GKA 534
DB 825 GETGPPGAGFPAGPPGADGQPKAGKDGQGEAGQKGDAGAPGQGPAGFCPQGTGVTGPK 884

QY 535 GERGVGPPGA-----VPGAKDGEAGAQPPGPA---GPAGERGEOQPGSPGPFQ 582
DB 885 GARGAQGPPGATGTFPGAAGRVGPPGANGNPPGAPGPPGAKDKGPKGVKRGDSGPPGRAGDP 944

QY 583 GLPGPAGPGGAGKPGEGVPGDLGAPGSPGAG 616

DB 945 GLQGPAGAPGKGEFGDDGSGLDGPPGPOQLAG 978

RESULT 9

CGB01S

collagen alpha 1(I) chain - bovine (tentative sequence) (fragments)

C:Species: Bos primigenius taurus (cattle)

C:Date: 24-Apr-1984 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004

C:Accession: A91193; A91229; A91387; A91201; A91200; A43048; A02853

R:Rauterberg, J.; Timpl, R.; Furthmayr, H.

Eur. J. Biochem. 27, 231-237, 1972

A:Title: Structural characterization of N-terminal antigenic determinants in calf and human

A:Reference number: A91193; MUID:7225334; PMID:4115172

A:Accession: A91193

A:Molecule type: protein

A:Residues: 1-19 <RAU>

A:Cross-references: UNIPROT:P02453

A:Experimental source: skin

A:Note: The epsilon carbon of Lys-9, by homology with the rat alpha 1(I) chain, is converted

R:Fietzek, P.P.; Kuehn, K.

Eur. J. Biochem. 52, 77-82, 1975

A:Title: The covalent structure of collagen: amino-acid sequence of the cyanogen-bromide

A:Reference number: A91229; MUID:76022320; PMID:1164916

A:Accession: A91229

A:Molecule type: protein

A:Residues: 20-145 <FIE>

A:Experimental source: skin

A:Note: Lys-103 is hydroxylated and binds glucosylgalactose

R:Fietzek, P.P.; Wendt, P.; Kell, I.; Kuehn, K.

FEBS Lett. 26, 74-76, 1972

A:Title: The covalent structure of collagen: amino acid sequence of alpha1-CB3 from calf

A:Reference number: A91387; MUID:73049499; PMID:4673951

A:Accession: A91387

A:Molecule type: protein

A:Residues: 146-294 <F12>

A:Experimental source: skin

R:Fietzek, P.P.; Rexrodt, F.W.; Hopper, K.E.; Kuehn, K.

Eur. J. Biochem. 38, 396-400, 1973

A:Title: The covalent structure of collagen. 2. The amino-acid sequence of alpha1-CB7 from

A:Reference number: A91211; MUID:74086118; PMID:4359390

A:Accession: A91211

A:Molecule type: protein

A:Residues: 295-562 <F13>

A:Experimental source: skin

R:Wendt, P.; Mark, K.V.D.; Rexrodt, F.; Kuehn, K.

Eur. J. Biochem. 30, 169-183, 1972

A:Title: The covalent structure of collagen. The amino-acid sequence of the 112 residues.

A:Reference number: A91201; MUID:73042276; PMID:4343808

A:Accession: A91201

A:Molecule type: protein

A:Residues: 563-675 <WEN>

A:Experimental source: skin

R:Fietzek, P.P.; Rexrodt, F.W.; Wendt, P.; Stark, M.; Kuehn, K.

Eur. J. Biochem. 30, 163-168, 1972

A:Title: The covalent structure of collagen. Amino acid sequence of peptide alpha1-CB6-C

A:Reference number: A91200; MUID:73042275; PMID:4343807

A:Accession: A91200

A:Molecule type: protein

A:Residues: 676-758 <F14>

A:Experimental source: skin

A:Note: Pro-726 is the only 3-hydroxyproline and the only hydroxylated proline in position

R:Rauterberg, J.; Fietzek, P.; Rexrodt, F.; Becker, U.; Stark, M.; Kuehn, K.

FEBS Lett. 21, 75-79, 1972

A:Title: The amino acid sequence of the carboxyterminal nonhelical cross link region of t

A:Reference number: A43048

A:Accession: A43048

A:Molecule type: protein

A:Residues: 759-779 <RA2>

A:Experimental source: skin

C:Comment: Lysines at positions 115, 124, 274, 346, 424, 496, 658, and 670 may be hydroxy

C:Comment: Prolines in the third position of the tripeptide repeating unit (G-X-Y) are hy

C;Comment: The order of the eight CNBR peptides in the alpha 1(I) chain of bovine skin d
9, 149, 268, and 217 residues.

C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyrrolidone carboxylic acid; trimer;
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 54.4%; Score 1898; DB 1; Length 779;
Best Local Similarity 53.0%; Pred. No. 9.6e-89;
Matches 384; Conservative 22; Mismatches 210; Indels 108; Gaps 10;

```

QY 1 GPPGPGPTGLPFPFGRGGPGS-----RGFGADGVAGPKPAGRGSPGAGPKGSP 54
DB 29 GPPGAGPQGFQGFPGFPGFPGSGPMGPRGPPGPKNGDDGAGKPGPRGPRGPPGQ 88
QY 55 GEARGPEAGLPGAKGLTSGSPGPDGKTTPPGPAGQDGRPPPPPPGARGQAGVMGFP 114
DB 89 GARGLPGTAGLPGMKHGRGSLDGAKDAGPAGPKPGSPGPGANGAPQMGPRGLPGFP 148
QY 115 GPKGAAGEPKKAGRGVPPRGAVCPAGKDGAGAGAQPPPPAGPAGRGEGQAGSPGFQ 174
DB 149 GPKGAAGEPKKAGRGVPPRGAVCPAGKDGAGAGAQPPPPAGPAGRGEGQAGSPGFQ 208
QY 175 GLPGPAGPPEAGKPGQGVPGDGLGAPGSPGAGE-----PGP-----212
DB 209 GLPGPAGPPEAGKPGQGVPGDGLGAPGSPGARGRGFPGERVVEGPPGPRGANGAP 268
QY 213 -----TGLPDPBRGPGSGRFPFGADGVAGPKPAGRGSGS-----248
DB 269 GNDGAKDAGAPAGPSQAGPLQPMGGERGAAGLPKPKGDRDAGPKGADGAPGKGYR 328
QY 249 -----PGPAGPKSPGAGRPGEAGLPGAKGLTSGSPGPDGKTTPPGAGQDGR- 299
DB 329 GLTGPTGPPGAPGDKGAGSPGAGTKGAPGDRGFPGPPGAPGAGPAGADGQPGAK 388
QY 300 -----PGPPGPGARGQAGVMGFPFGKGAAGEPKKAGRGVPPGPAVGPA 345
DB 389 GEPGDAGAKDAGPFPAGPAGPPGPIGNVAGPKGARGSPGATGFPGAAGRVGP 448
QY 346 GKDGAGAGPPGAPGAGE-----RGQAGPAGSPGFGPLGPPAGPGEAGKPGQGVPGDL 402
DB 449 GPGNAGPPGPPGAPGAKGSGKPRGTGAPRGVEGPPGPPGAPGAGADGAPAG 508
QY 403 GAPPGSPGAGEPGTGLPFPBRGPGSGRFPFGADGVAGPKPAGRGSPGAPG---K 459
DB 509 GTFGPQGIAGQGVVGLPQGRGEPFGLPFGSGEPKPGSGASGERPPGMPGPPGLA 568
QY 460 GSPGAGRPGEAGLPGAKGLTSGSPGPDGKTTPPGPAGQDGRFPFPFGARGQAGVM 519
DB 569 GPPGESGREGAPGAEGRDGSAGKAGDRGETGAPAGPAGPAGPAGPVPAGKSGDR 628
QY 520 GPPGPKGAAGEPKKAGRGVPPGPAVGPAKDG-----AGAQPPGP-----563
DB 629 GETGPAGTGPVGPAGAPGAPGQPRGKGTGZGZBRGKIKHGRFGSLQGPFGPPGSP 688
QY 564 -----AGPAGRGEGQAGSP---GFGPLGPPAGPPEAGKPGEGVPCDLGAPGS 612
DB 689 GEQPSGASGAPGPRGPPGASGPGKGLGGLPGLGIPGPRGRTGDTAGPAGPPGPP 748
QY 613 GPAG 616
DB 749 GPPG 752

```

RESULT 10

B40333

collagen alpha 1(II) chain precursor - African clawed frog

C;Species: Xenopus laevis (African clawed frog)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C;Accession: B40333

R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.

J. Cell Biol. 115, 565-575, 1991

A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em

A;Reference number: A40333; MUID:92011898; PMID:1918153

A;Accession: B40333

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1486 <SUA>

A;Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; GB:M63595

C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <WVC>
F;1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <PCC>

Query Match 54.2%; Score 1890.5; DB 1; Length 1486;
Best Local Similarity 54.6%; Pred. No. 3.6e-88;
Matches 374; Conservative 40; Mismatches 196; Indels 75; Gaps 10;

```

QY 1 GPPGPGPTGLPFPFGRGGPGSFGFPGADGVAGPKGPA-----GEGSGPGAPGPK 51
DB 269 GPPGQAGRGFPFGLFVKGHRGYPGLDGAKEAGAAAGAKGCGATGAGSGPMPGR 328
QY 52 GSPGAGRPGBAGLPGAKGLTSGSPGPDGKTTPPGAGQDGRPPGPPGARGQAGVM 111
DB 329 GLPGERGPGSSGAAGAR---GNDGLGFP---AGPPGVGAPGAPPGAPSGKEAGPT 382
QY 112 GPPGPKGAAGPBGKAGRGVPPGPAVCPAGKDGAGAGAQPPPPAGPAGRGEGQAGSP 171
DB 383 GARGPEGAQGRGESGTPGSPGAGASGNTDGLPGAKGSSGGPGIAGAPGFPGRGPP 442
QY 172 GFQGLPGPAGPPGAGKP-----GEOGVPCDLGAPGS---GPA-----GEPGPT 213
DB 443 GPQATGPLGPKGTGDPGVAGFKGEGQKGEISAGPQGAPGAPGEGKRGARGEGCAA 502
QY 214 GLPFPGERGPGSGRFPFGADGVAGPKPAGRGSGPPAGPKGSPGAGRPGEAGLPGAK 273
DB 503 GPNPGERGAPGNRGFPQDGLAGPKGAPGERGVPLGGLPKGNGDPGPRGEPGLPGAR 562
QY 274 GLTSGPSGPDGKTTPPGAGQDGRPPGPPGARGQAGVMGFPFGKGAAGEPFGKAGER 333
DB 563 GLTGRPDAGPQGVKVGSGASGEDRPPGPPGQARGQPVGMGFPFGKANGEPFGKAGEK 622
QY 334 GVPFPAGVAGPAGKDGAGAGQPPGPPAGPAGRGEGQAGSPGFGPLGPPAGPGEAGKP 393
DB 623 GLVGAPGLGLPKDGTGSGQPNPAGPAGRGEGQAGSPGFGPLGPPGPFGEAGKP 682
QY 394 GEQVPCDLGAPSGPAGE---PGPTGLPFPGERGPGSGRFPFGADGVAGPKGAPAGER 450
DB 683 GDQVPGEAGAGLVGPRGERGFPGERGSSGPGQLQGRGLPGTGTGDKGASGSGPN 742
QY 451 GSPGAPGKSPGEGAGRPGEAGLPGAKGLTSGSPGPDGKTTPPGAGQDGRPPGPP 510
DB 743 GAQFPGLQMPGERGAAGISGPKGDRGDTGERGPEGASGKDSRGLTGTGIPGPPGAPN 802
QY 511 GARGQAGVMGFPFGKGAAGEPFGKAGRGVPPGPAVGPAKGD-----GEA 555
DB 803 GEGSGSPGFPGLVGARGAPDRGENGPPGAPGAGPAGDQSGGLKDGQSGSGKGD 862
QY 556 GAQPPGPA-----GPAGRGEGQAGSPGFGQL-----PGPAGPP 591
DB 863 GAPFPQPSGAPGPQGTGVFGKARGAGAGPAGATGFPGAAGRVGTGPGNNGPFP 922
QY 592 GEAGKPGEGQVPCDLGAPGSPGAP 616
DB 923 GSAGKPGKVRGDAGPPGRAGDPG 947

```

RESULT 11

A40333

collagen alpha 1'(II) chain precursor - African clawed frog

C;Species: Xenopus laevis (African clawed frog)

C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004

C;Accession: A40333

R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.

J. Cell Biol. 115, 565-575, 1991

A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em

A:Reference number: A40333; MUID:92011898; PMID:1918153

A:Accession: A40333

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-1492 <SUA>

A:Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; GB:M63596

A:Note: This sequence is presented as substitutions relative to another sequence in a file they replace; the appropriate interpretation of the sequence figure was reconstructed
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:37-96/Domain: von Willebrand factor type C repeat homology <VMC>
F:1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 54.0%; Score 1883; DB 2; Length 1492;

Best Local Similarity 49.1%; Pred. No. 8.6e-88;

Matches 379; Conservative 46; Mismatches 185; Indels 162; Gaps 13;

QY 1 GPPGPGTGLPGPGRGGGSGRFFPCAD-----GVAGPKGAGERGSGPPAGPK 51

DB 272 GPPGPGGARGFTCTGLPGVKGHRGYPGLDGGKEAGAGAKGSCAGSAGAGAPGMPGR 331

QY 52 GSPGAGRPGEAGLPGAKLGTSGSPGPDGKTGTPPGAGQDGRPGPPPGPCARQAQGM 111

DB 332 GLPGRGRPGSSGAAGAR---GNDGLPGP---AGPPGVPAGAPGFFGAPGSGEAGPT 385

QY 112 GPPGPKGAAGEPKKAGERGVGP-----PGAVGAGKDGGEAGAGAGPPGAGPA 159

DB 386 GARGPEGAQGRGESGTPGSPGSGASGNPGTDTGPGAKGSSGASGIAGAPCFPGPRGPP 445

QY 160 GERGGQGP-----AGSPGFCGLPGAPGPGAGKPGQGVPGDLGAPGSGPAGEGPT 213

DB 446 GPGATGPLPGKETGDPGIAFGKEHGPKEIGSAGPQAGPPAGEGKRGARGEPAA 505

QY 214 GLPFPGRGGGSGRFFPCADGAVAGPKGAPGERSFGPAGPKSGPGEARFCEAGLPGAK 273

DB 506 GPLPFPGRGAPGNRGFFPCQDGLAGPKGAPGERGVPLGGPKGNGDGRPGERGLPGAR 565

QY 274 GLTSGFSGPDGKTGTPPGAGQDGRPPGPPGARGQAQVWVFPFGPKAAGEPKKAGER 333

DB 566 GLTGRPGDAGPKGVKPSGAAGEDRPGPPGQARGQGVWVFPFGPKGANGEPKAGEK 625

QY 334 GVPGPGVAGVAGKDGAGAGACPPGAPGAGERGQGPAGSGPFGGLPGAPGPEACKP 393

DB 626 GLGGAPGLRGLPGKDGETGAQGNPGAPGAGERGQGPFGGFGGLPGPPSGPGECKP 685

QY 394 GQGVPGDL-----GAPGSPGA 411

DB 686 GQGVPGEAGAGLVPRGERGFPGERGSSPGQGLQSGRLPGTPTDGPKGATGSPGN 745

QY 412 GEPGPTGLPGPPGERG-----GP-----GSRFFPCADGAVAGPKGPA 447

DB 746 CAQGPPLQGMPCGERGAAGISGPKDRDGTGKGPAGAPGKDGSRGLTGLPLGPPGAPGN 805

QY 448 GERGSFGAPG-----KGSFGAERPGEA 471

DB 806 GKSGSGSPGPIGVARGAPGDRGENGPPGAPGPPGSDGQAGLKGDSGSKGDA 865

QY 472 GLPGAKGLTSGSPGPDGKTGP-----PGPAGQDGRPGP---PGPGAR 513

DB 866 GAPGQPSGAPGPOGTGVNGPKARGAQGPAGATGFPGAAGRVTGTPGNPNTGPPPP 925

QY 514 QAQVWVFPFGPKAAGEPKKAGERGVPPGPPGAVGAPGAKDGEAGAGAPGPPGAPGERGEQ 573

DB 926 GSAGKEGPKGVGDAGPTGRAGDGLQGPAGAPGKEGPEGDPGSDGPPGQSLSGNR 985

QY 574 GPAGSP-----GFOGLPGAPGPEACKP-----GQGVPGDLGAPGSPGAG 616

DB 986 GIVGLPQGRGERGFPGLPFPSPGPKQGGPSGSDRPPGPPGVPPGLTGPAG 1037

RESULT 12

S59856

collagen alpha 1(III) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 09-Jul-2004

C:Accession: S59856; S62120; S16373

R:Toman, P.D.; de Crombrughe, B.

Gene 147, 161-168, 1994

A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA

A:Reference number: S59856; MUID:95011609; PMID:7956795

A:Accession: S59856

A:Molecule type: DNA

A:Residues: 1-1464 <TOM>

A:Cross-references: UNIPROT:P08121; EMBL:X52046

R:Toman, D.

submitted to the EMBL Data Library, November 1994

A:Reference number: S62120

A:Accession: S62120

A:Molecule type: DNA

A:Residues: 1-866,'G',868-1464 <TOM>

A:Cross-references: EMBL:X52046; NID:G575321; PIDN:CAA36279.1; PID:G575322

R:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

Biochim. Biophys. Acta 1089, 241-243, 1991

A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.

A:Reference number: S16176; MUID:91274355; PMID:2054384

A:Accession: S16373

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1442-1464 <MET>

A:Cross-references: EMBL:X57983; NID:G50476; PIDN:CAA1048.1; PID:G50477

C:Genetics:

S58/3; 673/3; 706/3; 742/3; 760/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 976/3;

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C:Keywords: coiled coil; extracellular matrix

F:1-24/Domain: signal sequence #status predicted <SIG>

F:25-154/Domain: propeptide #status predicted <PRO>

F:32-92/Domain: von Willebrand factor type C repeat homology <VMC>

F:155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>

F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 53.3%; Score 1857.5; DB 2; Length 1464;

Best Local Similarity 53.4%; Pred. No. 1.6e-86;

Matches 369; Conservative 39; Mismatches 190; Indels 93; Gaps 12;

QY 1 GPPGPGTGLPGPGRGGGSGRFFPCADGAVAGPKGAPGERSGPPAGPKSGPGEAGRP 60

DB 467 GSPGECANGLPAGAAAGRSGSPFRGAPGNIPGKEGPPGERGPGPAGRGVAGEPRD 526

QY 61 GEAGLPGAKGLTSGSPGPDGKTGTPPGAGQDGRPPGPPGARGQAQVWVFPFGPKGAA 120

DB 527 GTPGPGIRGMPGSPGPGGNDCKPGPPGSGQGESRPPGPPGPPGPPGPPGPPGPPG 586

QY 121 GEPKAGERGVPPGPPGAVGVPAGKDGGEAGAGAQPPGAPGAPGERSGPPGPPGPPGPPG 180

DB 587 GAPKNGERGGPGGPGPLPGPAGKNGETGPPGPPGTPGAGDKGSDGPPGPPGPPGPPG 646

QY 181 GPPGKAGKPGBOGVPLDLGAPGSPGAPGPTGLPGPPGPPGPPGPPGPPGPPGPPGPPG 240

DB 647 GPPGKNGKPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 700

QY 241 GPAGERSGPPGAPGPKSGPGEAGRPGEAGLPGAKGLTSGSPGPPGPPGPPGPPGPPGPPG 300

DB 701 GP---EGKGPAGPPGPPGASGSPGLOQMPGER---GGPSPGPPGPPGPPGPPGPPGPPG 754

QY 301 GPPGPPGARGQAQVWVFPFGPKAAGFPKAGERGVPVPPGAVGAPGAKDGEAGAGAGPPGPPG 360

DB 755 GKDG---RGPAFPIGPPG---AGQPDKGEKGGSPGLPGIAGPRGPPGPPGPPGPPGPPG 808

QY 361 -----GPAGERGQGPAGSPGFOGLPGAPGPPGE---AGKPGQGVPPDGLGAPGPPS 408

DB 809 GPPGAPGQNGEPGAKGERGAPGKEGEGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 868

QY 409 GPAGERGPTGLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 468

DB 869 GTAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 928

Db 1032 GPPG 1035

RESULT 14
CGHU7L

collagen alpha 1(III) chain precursor - human
N:Alternate names: procollagen alpha 1(III) chain
C:Species: Homo sapiens (man)
C:Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 09-Jul-2004
C:Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90399
R:Prockop, D.J.
submitted to the EMBL Data Library, February 1989
A:Reference number: S05272
A:Accession: S05272
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1240, 'Y', 1242-1466 <PRC>
A:Cross-references: UNIPROT:P02461; EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
R:Ala-Kokko, L.; Kontusari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.
Biochem. J. 260, 509-516, 1989
A:Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human
enriches.
A:Reference number: S04642; MUID:89350839; PMID:2764886
A:Accession: S04642
A:Molecule type: mRNA
A:Residues: 1-1196 <ALA>
A:Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
A:Note: the complete sequence is not shown
R:Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.
Gene 78, 235-265, 1989
A:Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene
A:Reference number: PE0011; MUID:89378752; PMID:2777083
A:Accession: PE0011
A:Molecule type: DNA
A:Residues: 1-176 <BEN>
A:Cross-references: GB:M26939; NID:g180813; PIDN:AA52040.1; PID:gi180814
R:Toman, P.D.; Ricca, G.A.; de Crombrughe, B.
Nucleic Acids Res. 16, 7201, 1988
A:Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre
A:Reference number: S01726; MUID:88303360; PMID:3405773
A:Accession: S01726
A:Molecule type: mRNA
A:Residues: 1-170 <TOM>
A:Cross-references: EMBL:X07240; NID:g30060; PIDN:CAA30229.1; PID:g30061
A:Note: the authors translated the codon CAG for residue 154 as His
R:Janeczko, R.A.; Ramirez, F.
Nucleic Acids Res. 17, 6742, 1989
A:Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.
A:Reference number: S04887; MUID:89386015; PMID:2780304
A:Accession: S04887
A:Molecule type: mRNA
A:Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634, 'A'
A:Cross-references: EMBL:X15332; NID:g29545; PIDN:CAA33387.1; PID:g930045
A:Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
R:Seyer, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A:Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A:Reference number: A90399; MUID:77134724; PMID:557335
A:Accession: A90399
A:Molecule type: protein
A:Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>
A:Experimental source: liver
A:Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galact
R:Seyer, J.M.
submitted to the Atlas, December 1977
A:Reference number: A94562
A:Accession: A94562
A:Molecule type: protein
A:Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>
A:Experimental source: liver
A:Note: author submitted corrections to A90399
R:Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Bvers, P.H.

Am. J. Hum. Genet. 53, 62-70, 1993

A;Title: Parental somatic and germ-line mosaicism for a multixon deletion with unusual splicing.

A;Reference number: I51868; MUID:93304430; PMID:8317500

A;Accession: I51868

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 186-194 <MIL>

A;Cross-references: GB:S62925; NID:g386425; PIDN:AAD13937.1; PID:g4261637

R;Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.

Biochem. J. 311, 939-943, 1995

A;Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3A1 gene.

A;Reference number: S59511; MUID:96067614; PMID:7487954

A;Accession: S59511

A;Status: translation not shown

A;Molecule type: mRNA

A;Residues: 302-423 <CHI>

A;Cross-references: GB:S79877; NID:gl195576; PIDN:AAB35615.1; PID:gl195577

R;Seyer, J.M.; Kang, A.H.

Biochemistry 17, 3404-3411, 1978

A;Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr peptides.

A;Reference number: A90414; MUID:79000343; PMID:687591

A;Accession: A90414

A;Molecule type: protein

A;Residues: 339-675, N', 677-727 <SEY3>

A;Experimental source: liver

R;Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.

J. Biol. Chem. 266, 5256-5259, 1991

A;Title: G to T transversion at position +5 of a splice donor site causes skipping of the exon 17 in the COL3A1 gene.

A;Reference number: I55349; MUID:91161621; PMID:1672129

A;Accession: I55349

A;Status: translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 537-605 <LEE>

A;Cross-references: GB:M59312; NID:gl180815; PIDN:AAA52041.1; PID:gl180816

R;Seyer, J.M.; Mainardi, C.; Kang, A.H.

Biochemistry 19, 1583-1589, 1980

A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from type I procollagen.

A;Reference number: A90438; MUID:80198282; PMID:6246925

A;Accession: A90438

A;Molecule type: protein

A;Residues: 728-895, 'A', 897-964 <SEY4>

A;Experimental source: liver

R;Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan, J.

J. Biol. Chem. 265, 17070-17077, 1990

A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping and results in a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syndrome.

A;Reference number: A38303; MUID:91009133; PMID:2145268

A;Accession: A38303

A;Molecule type: mRNA

A;Residues: 861-1015 <COL>

A;Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:gl180878; PIDN:AAB59383.1; PID:g180878

A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syndrome.

R;Mankoo, B.S.; Dalglish, R.

Nucleic Acids Res. 16, 2337, 1988

A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.

A;Reference number: S02119; MUID:88189827; PMID:3357782

A;Accession: S02119

A;Status: translation not shown

A;Molecule type: mRNA

A;Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>

A;Cross-references: EMBL:X06700; NID:g30053; PIDN:CAA29886.1; PID:g30054

R;Seyer, J.M.; Kang, A.H.

Biochemistry 20, 2621-2627, 1981

A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from type I procollagen.

A;Reference number: A90446; MUID:81208139; PMID:7016180

A;Accession: A90446

A;Molecule type: protein

A;Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-1159, 'S', 1160-1161, 'S', 1162-1163, 'S', 1164-1165, 'S', 1166-1167, 'S', 1168-1169, 'S', 1170-1171, 'S', 1172-1173, 'S', 1174-1175, 'S', 1176-1177, 'S', 1178-1179, 'S', 1180-1181, 'S', 1182-1183, 'S', 1184-1185, 'S', 1186-1187, 'S', 1188-1189, 'S', 1190-1191, 'S', 1192-1193, 'S', 1194-1195, 'S', 1196-1197, 'S', 1198-1199, 'S', 1200-1201, 'S', 1202-1203, 'S', 1204-1205, 'S', 1206-1207, 'S', 1208-1209, 'S', 1210-1211, 'S', 1212-1213, 'S', 1214-1215, 'S', 1216-1217, 'S', 1218-1219, 'S', 1220-1221, 'S', 1222-1223, 'S', 1224-1225, 'S', 1226-1227, 'S', 1228-1229, 'S', 1230-1231, 'S', 1232-1233, 'S', 1234-1235, 'S', 1236-1237, 'S', 1238-1239, 'S', 1240-1241, 'S', 1242-1243, 'S', 1244-1245, 'S', 1246-1247, 'S', 1248-1249, 'S', 1250-1251, 'S', 1252-1253, 'S', 1254-1255, 'S', 1256-1257, 'S', 1258-1259, 'S', 1260-1261, 'S', 1262-1263, 'S', 1264-1265, 'S', 1266-1267, 'S', 1268-1269, 'S', 1270-1271, 'S', 1272-1273, 'S', 1274-1275, 'S', 1276-1277, 'S', 1278-1279, 'S', 1280-1281, 'S', 1282-1283, 'S', 1284-1285, 'S', 1286-1287, 'S', 1288-1289, 'S', 1290-1291, 'S', 1292-1293, 'S', 1294-1295, 'S', 1296-1297, 'S', 1298-1299, 'S', 1300-1301, 'S', 1302-1303, 'S', 1304-1305, 'S', 1306-1307, 'S', 1308-1309, 'S', 1310-1311, 'S', 1312-1313, 'S', 1314-1315, 'S', 1316-1317, 'S', 1318-1319, 'S', 1320-1321, 'S', 1322-1323, 'S', 1324-1325, 'S', 1326-1327, 'S', 1328-1329, 'S', 1330-1331, 'S', 1332-1333, 'S', 1334-1335, 'S', 1336-1337, 'S', 1338-1339, 'S', 1340-1341, 'S', 1342-1343, 'S', 1344-1345, 'S', 1346-1347, 'S', 1348-1349, 'S', 1350-1351, 'S', 1352-1353, 'S', 1354-1355, 'S', 1356-1357, 'S', 1358-1359, 'S', 1360-1361, 'S', 1362-1363, 'S', 1364-1365, 'S', 1366-1367, 'S', 1368-1369, 'S', 1370-1371, 'S', 1372-1373, 'S', 1374-1375, 'S', 1376-1377, 'S', 1378-1379, 'S', 1380-1381, 'S', 1382-1383, 'S', 1384-1385, 'S', 1386-1387, 'S', 1388-1389, 'S', 1390-1391, 'S', 1392-1393, 'S', 1394-1395, 'S', 1396-1397, 'S', 1398-1399, 'S', 1400-1401, 'S', 1402-1403, 'S', 1404-1405, 'S', 1406-1407, 'S', 1408-1409, 'S', 1410-1411, 'S', 1412-1413, 'S', 1414-1415, 'S', 1416-1417, 'S', 1418-1419, 'S', 1420-1421, 'S', 1422-1423, 'S', 1424-1425, 'S', 1426-1427, 'S', 1428-1429, 'S', 1430-1431, 'S', 1432-1433, 'S', 1434-1435, 'S', 1436-1437, 'S', 1438-1439, 'S', 1440-1441, 'S', 1442-1443, 'S', 1444-1445, 'S', 1446-1447, 'S', 1448-1449, 'S', 1450-1451, 'S', 1452-1453, 'S', 1454-1455, 'S', 1456-1457, 'S', 1458-1459, 'S', 1460-1461, 'S', 1462-1463, 'S', 1464-1465, 'S', 1466-1467, 'S', 1468-1469, 'S', 1470-1471, 'S', 1472-1473, 'S', 1474-1475, 'S', 1476-1477, 'S', 1478-1479, 'S', 1480-

Query March 52.8%; Score 1840; DB 2; Length 886;
 Best Local Similarity 49.8%; Pred. No. 8.4e-86;
 Matches 362; Conservative 44; Mismatches 195; Indels 126; Gaps 11;
 A:Gene: COL3A1
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 F:30-90/Domain: von Willebrand factor type C repeat homology <VWC>
 C:Genetics:
 A:Cross-references: UNIPROT:P12105; EMBL:U07973; NID:G520454; PIDN:AAA63407.1; PID:G5374;
 C:Comments:

```

QY 1 GPPGPGTGLPGRGPGSGRF-----PGADGVAGPKGPAGERGSGPAGPKGSGCEA 57
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
173 GPPGPGSGPGPGHAGPGSGNGYQPPGPGPGPGPGPGPGPGPGPGPGPGPGPG 232
QY 58 GRP-----GEAGLPKAKGLTSGSP 78
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
233 GPPGNGDRG1PGLPGHKHGMPCMGKMGARGFDGDKGAKGDSGAPGPKGEAGQPGAN 292
QY 79 GPDGKTGPPGAGQDGRPGPPGARGO-----AGVMGPPGPKAAAGEPKKAGERGVP 132
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
293 GSPGPGPGGTGERGPGNPGGPGAGKGDAGTAGPLGPPGPGTAGTGGPGSGFKGEA 352
QY 133 GPPGAVGAGKDGAGAGAGGPPGAGPAGERGEGQAGSGPGQLGPA-----180
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
353 GPPGAGASGNPGERGEPGQAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 412
QY 181 ---GPPGAGKPGEGQVP-----GDLGAPGSPGPA-----CEP 210
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
413 GLPGPPTSGNPGAKGTGEPGKNGAKGDPGPKGERGNGTFCARGPPGEEKRGANGEP 472
QY 211 GPTGLPGPPGERGPGSGRFPAGDGVAGPKGPAGERGSGPAGPKGSGPGEAGRPGEAGLP 270
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
473 GQGVGPGTPGERGSGFRGLPGSNGLPGEGKGPAGERGSGPPGPGSGPAGDRQDGGPLP 532
QY 271 GAKGLTSGSGSPGPKGTGPPGAGQDGRPPGPPGPPGPPGPPGPPGPPGPPGPPG 330
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
533 GMRGLPGIPGSGSGKPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 592
QY 331 GERGVGPPGAVGAGKDGAGAGGPPGAGPAGERGEGQAGSGPGQGLPGAGPPGGEA 390
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
593 GERGGPPGTPGPAKNGDVGLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 652
QY 391 GKPGGQGVGDLGAPGSPGPP-----AGEPGTGLPGPPGERGSGRGPFGADGVAGPK 444
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
653 GKPGGPGKDIIGGPGPKGNGIIPGERGPPGPPGPPGPPGPPGPPGPPGPPGPP 697
QY 445 GPAGERGSGPAGPKGSGPGEAGRPGEAGLPKAKGLTSGP-----GSPGPDGKTGPPGPA 498
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
698 GPAGSEGAKGPPGPGAGPGTGLPGQMPGERGASGPGPKGDKGEPGKGADGLPGAR 757
QY 499 GQDGR-----PGPPGPPGARGOAGVMGPPGPKGAAGBPGKAGERGVPGP---PGAVGPA 549
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
758 GERGNVPIGPPGPPGPPGDKGETGAPAGPAGSGRGGGERGEGQLPGPAGFPAGQN 817
QY 550 GKDGEAGAGQPPGPPGAGERGEGQAGSGPGQGLPGPAGPPEAGKPGEGQGVPGDLGAP 609
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
818 GEPGKGERGPPGLRGEAGPPCAAGPQGGPGAGPPGPPGQGVKGERGSGPGGAGFPGAR 877
QY 610 GPPGPG 616
Db | | | | |
878 GPPGPPG 884

```

Search completed: June 17, 2005, 15:15:17
Job time : 26.1205 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:17:54 ; Search time 91.5221 Seconds
(without alignments)
3452.202 Million cell updates/sec

Title: US-10-658-989A-2

Perfect score: 3488

Sequence: 1 GPPGEPGTGLPGRPGRGG.....GEQGVPGDLGAPGSPGAGG 617

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	2237.5	64.1	1464	1 CALL_HUMAN	P02452 homo sapien
2	2233.5	64.0	1461	2 O76045	O76045 homo sapien
3	2233.5	64.0	1464	2 Q8N473	Q8N473 homo sapien
4	2231.5	64.0	1460	1 CALL_CANFA	Q9X8J7 canis famil
5	2222	63.7	1069	2 Q6LAN8	Q6LAN8 homo sapien
6	2215.5	63.5	1453	2 Q63079	Q63079 rattus norv
7	2205.5	63.2	1453	1 CALL_MOUSE	P11087 mus musculu
8	2205.5	63.2	1453	2 Q810J9	Q810J9 mus musculu
9	2176.5	62.4	1225	2 Q6PCL3	Q6PCL3 mus musculu
10	2151.5	61.7	1453	1 CALL_CHICK	P02457 gallus gall
11	2043.5	58.6	1450	2 O9VIB4	O9VIB4 cynops pyrr
12	2042.5	58.6	1445	2 Q93251	Q93251 rana catesb
13	2037.5	58.4	1449	2 Q640B2	Q640B2 xenopus tro
14	2029	58.2	1449	2 Q802B5	Q802B5 xenopus lae
15	2010.5	57.6	671	1 CALL_RAT	P02454 rattus norv
16	1976.5	56.7	1447	2 Q9TB91	Q9TB91 xenopus lae
17	1953	56.0	1160	2 Q14046	Q14046 homo sapien
18	1953	56.0	1487	2 Q14047	Q14047 homo sapien
19	1952	56.0	1487	2 Q17753	Q17753 canis famil
20	1944	55.7	1418	2 Q82396	Q82396 equus cabal
21	1943	55.7	1418	1 CALL_HUMAN	P02458 homo sapien
22	1942	55.7	1420	2 Q90W37	Q90W37 gallus gall
23	1937	55.5	1269	2 Q7T2Z7	Q7T2Z7 gallus gall
24	1936	55.5	1442	2 Q62031	Q62031 mus musculu
25	1936	55.5	1442	2 Q62033	Q62033 mus musculu
26	1936	55.5	1459	1 CALL_MOUSE	P28481 mus musculu
27	1936	55.5	1459	2 Q62032	Q62032 mus musculu
28	1927	55.2	1419	2 Q80VY3	Q80VY3 mus musculu
29	1927	55.2	1419	2 Q80X38	Q80X38 mus musculu
30	1927	55.2	1487	2 Q641K3	Q641K3 mus musculu
31	1923	55.1	1447	2 Q6U1J5	Q6U1J5 brachydanio

32	1921	55.1	1419	2	Q63123	Q63123 rattus norv
33	1914	54.9	1447	2	Q6P4U1	Q6P4U1 brachydanio
34	1898	54.4	779	1	CALL_BOVIN	P02453 bos taurus
35	1892	54.2	1492	2	Q6P4Z2	Q6P4Z2 xenopus tro
36	1890.5	54.2	1486	2	Q91717	Q91717 xenopus lae
37	1888.5	54.1	1486	2	Q7ZTI6	Q7ZTI6 xenopus lae
38	1883	54.0	1418	2	Q9W7R9	Q9W7R9 cynops pyrr
39	1880	53.9	1491	2	Q91718	Q91718 xenopus lae
40	1880	53.9	1491	2	Q7ZTM4	Q7ZTM4 xenopus lae
41	1873	53.7	1262	1	CALL_CHICK	P12105 gallus gall
42	1873	53.7	1449	2	Q910C0	Q910C0 oncorhynch
43	1860	53.3	1449	2	Q6PEI9	Q6PEI9 brachydanio
44	1858	53.3	1449	2	Q6NZ15	Q6NZ15 brachydanio
45	1856	53.2	1049	1	CALL_BOVIN	P04258 bos taurus

ALIGNMENTS

RESULT 1
CALL_HUMAN
ID CALL_HUMAN STANDARD; PRT; 1464 AA.
AC P02452; P78441; Q13896; Q13902; Q13903; Q14037; Q14992; Q15176;
AC Q15201; Q16050; Q7KZ30; Q7KZ34; Q8IVI5; Q9UML6; Q9UNW7;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT SER-1434.
RA Dalgleish R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-589 FROM N.A.
RX MEDLINE=88329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
RA D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez P.,
RA Pretorius P.J.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene
RT (COL1A1).";
RL Gene 67:105-115(1988).
RN [3]
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockup D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen";
RL Biochem. J. 253:919-922(1988).
RN [4]
RP SEQUENCE OF 1-181 FROM N.A.
RX MEDLINE=84270697; PubMed=6462220;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA Myers J., Williams C., Ramirez P.;
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT conservation of a pattern of introns and exons.";
RL Nature 310:337-340(1984).
RN [5]
RP SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.
RC TISSUE=Skin;
RX MEDLINE=71038625; PubMed=5529814;
RA Click E.M., Bornstein P.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 1 and alpha 2 chains of human skin collagen.";
RL Biochemistry 9:4699-4706(1970).
RN [6]
RP SEQUENCE OF 425-1464 FROM N.A.
RX MEDLINE=84080385; PubMed=6689127;

RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
 RA Prockop D.J.;
 RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
 RT of alpha 1 chain of human type I procollagen. Statistical evaluation
 RT of structures that are conserved during evolution.";
 RL Biochemistry 22:5213-5223(1983).
 RN [7]
 RP SEQUENCE OF 472-607 FROM N.A.
 RX PubMed=2981843;
 RA Chu M.-L., Gargiulo V., Williams C.J., Ramirez F.;
 RT "Multixon deletion in an osteogenesis imperfecta variant with
 RT increased type III collagen mRNA";
 RL J. Biol. Chem. 260:691-694(1985).
 RN [8]
 RP SEQUENCE OF 488-625 FROM N.A.
 RX PubMed=3857621;
 RA Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gelinas R.E.;
 RT "Intron-mediated recombination may cause a deletion in an alpha 1 type
 RT I collagen chain in a lethal form of osteogenesis imperfecta";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874(1985).
 RN [9]
 RP SEQUENCE OF 1179-1464 FROM N.A., VARIANTS OF-II HIS-1277; ARG-1388 AND
 RP 1337-GLU-TYR-1338 DEL, AND VARIANT SER-1434.
 RX MEDLINE=93352646; PubMed=8349697;
 RA Chessler S.D., Wallis G.A., Byers P.H.;
 RT "Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I)
 RT chain of type I collagen result in defective chain association and
 RT produce lethal osteogenesis imperfecta";
 RL J. Biol. Chem. 268:18218-18225(1993).
 RN [10]
 RP SEQUENCE OF 1229-1454 FROM N.A.
 RC TISSUE=Bone;
 RX MEDLINE=88124208; PubMed=3340531;
 RA Maekelae J.K., Raassina M., Virta A., Vuorio E.;
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
 RT domain";
 RL Nucleic Acids Res. 16:349-349(1988).
 RN [11]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=85130970; PubMed=2857713;
 RA Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;
 RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
 RT Promoter structure, Alu repeats, and polymorphic transcripts";
 RL J. Biol. Chem. 260:2315-2320(1985).
 RN [12]
 RP SEQUENCE OF 1-34 FROM N.A.
 RX MEDLINE=88097389; PubMed=3480516;
 RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
 RT "Regulatory elements in the first intron contribute to transcriptional
 RT control of the human alpha 1(I) collagen gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
 RN [13]
 RP SEQUENCE OF 1-44 FROM N.A.
 RX MEDLINE=88033098; PubMed=2822714;
 RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
 RA de Wet W.;
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)
 RT collagen gene enhance transcription";
 RL J. Biol. Chem. 262:15151-15157(1987).
 RN [14]
 RP SEQUENCE OF 33-52.
 RX PubMed=2318855;
 RA Wirtz M.K., Keene D.R., Hori H., Glanville R.W., Steinmann B.,
 RA Rao V.H., Hollister D.W.;
 RT "In vivo and in vitro noncovalent association of excised alpha 1 (I)
 RT amino-terminal propeptides with mutant pro alpha 2(I) collagen chains
 RT in native mutant collagen in a case of Ehlers-Danlos syndrome, type
 RT VII";
 RL J. Biol. Chem. 265:6312-6317(1990).
 RN [15]
 RP SEQUENCE OF 156-183 FROM N.A.
 RX PubMed=2767050;
 RA Weil D., D'Alessio M., Ramirez F., de Wet W., Cole W.G., Chan D.,

RA Bateman J.F.;
 RT "A base substitution in the exon of a collagen gene causes alternative
 RT splicing and generates a structurally abnormal polypeptide in a
 RT patient with Ehlers-Danlos syndrome type VII";
 RL EMBO J. 8:1705-1710(1989).
 RN [16]
 RP SEQUENCE OF 175-187 AND 274-289.
 RX PubMed=2169412;
 RA Baetge B., Notbohm H., Diebold J., Lehmann H., Bodo M., Deutzmann R.,
 RA Muller P.K.;
 RT "A critical crosslink region in human-bone-derived collagen type I.
 RT Specific cleavage site at residue Leu95";
 RL Eur. J. Biochem. 192:153-159(1990).
 RN [17]
 RP SEQUENCE OF 263-268.
 RC TISSUE=Skin;
 RX MEDLINE=71001508; PubMed=4319110;
 RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
 RT "A comparative study of glycopeptides derived from selected vertebrate
 RT collagens: A possible role of the carbohydrate in fibril formation";
 RL J. Biol. Chem. 245:5042-5048(1970).
 RN [18]
 RP SEQUENCE OF 281-302; 402-420; 823-842; 924-944; 1026-1045 AND
 RP 1143-1162 FROM N.A.
 RX PubMed=2374517;
 RA Labhard M.E., Hollister D.W.;
 RT "Segmental amplification of the entire helical and telopeptide regions
 RT of the cDNA for human alpha 1 (I) collagen";
 RL Matrix 10:124-130(1990).
 RN [19]
 RP SEQUENCE OF 425-490; 965-1024; 999-1039 AND 1453-1464 FROM N.A.
 RX MEDLINE=83064528; PubMed=6183642;
 RA Chu M.-L., Myers J.C., Bernard M.P., Ding J.-F., Ramirez F.;
 RT "Cloning and characterization of five overlapping cDNAs specific for
 RT the human pro alpha 1(I) collagen chain";
 RL Nucleic Acids Res. 10:5925-5934(1982).
 RN [20]
 RP SEQUENCE OF 710-745 FROM N.A., AND VARIANT OI-II ARG-728.
 RX PubMed=2339700;
 RA Wallis G.A., Starman B.J., Zinn A.B., Byers P.H.;
 RT "Variable expression of osteogenesis imperfecta in a nuclear family is
 RT explained by somatic mosaicism for a lethal point mutation in the
 RT alpha 1(I) gene (COL1A1) of type I collagen in a parent";
 RL Am. J. Hum. Genet. 46:1034-1040(1990).
 RN [21]
 RP SEQUENCE OF 746-781 FROM N.A., AND VARIANT OI-III SER-767.
 RX MEDLINE=95187161; PubMed=7881420;
 RA Forlino A., Zolezzi F., Valli M., Pignatti P.F., Cetta G.,
 RA Brunelli P.C., Mottes M.;
 RT "Severe (type III) osteogenesis imperfecta due to glycine
 RT substitutions in the central domain of the collagen triple helix";
 RL Hum. Mol. Genet. 3:2201-2206(1994).
 RN [22]
 RP SEQUENCE OF 1187-1220 FROM N.A., AND VARIANT CYS-1195.
 RX PubMed=3170557;
 RA Cohn D.H., Apone S., Eyre D.R., Starman B.J., Andreassen P.,
 RA Charbonneau H., Nicholls A.C., Pope F.M., Byers P.H.;
 RT "Substitution of cysteine for glycine within the carboxyl-terminal
 RT telopeptide of the alpha 1 chain of type I collagen produces mild
 RT osteogenesis imperfecta";
 RL J. Biol. Chem. 263:14605-14607(1988).
 RN [23]
 RP SEQUENCE OF 1440-1464 FROM N.A.
 RX MEDLINE=90110490; PubMed=2295701;
 RA Willing M.C., Cohn D.H., Byers P.H.;
 RT "Frameshift mutation near the 3' end of the COL1A1 gene of type I
 RT collagen predicts an elongated pro alpha 1(I) chain and results in
 RT osteogenesis imperfecta type I";
 RL J. Clin. Invest. 85:282-290(1990).
 RN [24]
 RP SEQUENCE OF 1454-1464 FROM N.A.
 RX MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
 RA Maatta A., Bornstein P., Penttinen R.P.;

RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";

Query Match 64.1%; Score 2237.5; DB 1; Length 1464;
Best Local Similarity 60.2%; Pred. No. 3.6e-81;
Matches 441; Conservative 23; Mismatches 152; Indels 117; Gaps 12;

QY 1 GPP-----GEGPTGLPGRGGRGPGGSRGPGGADGVAGPKGAGERGSPGA 48
DB 458 GPPGAGBEGKRGARGEGPTGLPGRGGRGPGGSRGPGGADGVAGPKGAGERGSPGA 517
QY 49 GPKGSPGAGRGAGLPCAKGLTSGSGSPGDKTGPAGQDGRGPPGPPGARGQA 108
DB 518 GPKGSPGAGRGAGLPCAKGLTSGSGSPGDKTGPAGQDGRGPPGPPGARGQA 577
QY 109 GVMGPPGKGAAGEPGKAGRGVPPGAVGAGKDGAGAGQGGPPGAGPAGERGEQGA 168
DB 578 GVMGPPGKGAAGEPGKAGRGVPPGAVGAGKDGAGAGQGGPPGAGPAGERGEQGA 637
QY 169 GSPGQGLPGAGPPGAGPGEAGKPGEGQVPGDLGAPGSGPAGE-----PGP----- 212
DB 638 GSPGQGLPGAGPPGAGPGEAGKPGEGQVPGDLGAPGSGPAGE-----PGP----- 697
QY 213 -----TGLPGPPGGRGGRGPGGSRGPGGADGVAGPKGAGERGSPGAPGKGP 258
DB 698 GAGAPGNDGAKGADGAGAPGSGAGLQGMGPRGGAAGLPGPKGDRGDGPKGADGSP 757
QY 259 GE-----AGRPGGAGLPGAK-----GLTSGSPGSPGDKTGPAGQDGRP 300
DB 758 GKGVGRGLTGTGTPGPGAGAGKDGESGSPGAGTGGAGAPGDRGPPGPGAGFAGPP 817
QY 301 GPPGPPGARGQ-----AGVMGPPGKGAAGEPGKAGRGVPPGPA----- 341
DB 818 GADGQGAKGEGPDGAGAKGDAGPPGAGPPGPGPIGNVGAPGAKGARGAGPPGATGFP 877
QY 342 -----VGPAGKDGAGAGGPPGAGP-----AGRGQGAGSPGFGGLGPPAGPGEAGKP 393
DB 878 GAAGRVGPPGSGNAGPPGPPGAKGEGKGRGETGAGRGVPPGPPGAGKGP 937
QY 394 GPGVPGDLGAPGSPGAGEPGTGLPGRGGRGPGGSRGPGGADGVAGPKGAGERGSP 453
DB 938 GADGAGAGTGTGPOGLAGQGVGLPQGRGEGFGLPGSGEGPKQSPGASGERGPP 997
QY 454 GPAGP---KGSPPGAGRGAGLPAKGLTSGSPGSPGDKTGPAGQDGRGPPGPP 510
DB 998 GPMGPPGLAGPPGSGREGAPCAEGSPCRDGSAGKGRGETGAGPPGPPGAPGAPGV 1057
QY 511 GARGQAGVMGPPGKGAAGEPGKAGRGVPPGAVGAGKDG-----AGAQ 558
DB 1058 GPAGKSGDRGETGAPGAPGVGPPGAGPAGPQGPGRGDKGETGEGQDGRGKGRGFSGLQ 1117
QY 559 GPPGP-----AGPAGRGEGQGPAGSP---GFGGLPGAPGPPGAGKPGEGQGV 603
DB 1118 GPPGPPGSPGEGPGSAGSPAGPGRPPGAGAPKDGGLNGLPPIGPPGPPGRTGDAGPV 1177
QY 604 GDLGAPGSPGAPG 616
DB 1178 GPPGPPGPPGPPG 1190

RESULT 2

ID 076045 PRELIMINARY; PRT; 1461 AA.
AC DT 01-NOV-1998 (TREMELrel. 08, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DE 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Pro alpha 1(I) collagen.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

[1] SEQUENCE FROM N.A.
RP MEDLINE=85130970; PubMed=2857713;
RX Chu M.L., de Wet W., Bernard M., Ramirez F.;
RA "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
[2] SEQUENCE FROM N.A.
RP MEDLINE=88329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
RX D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RA "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene.";
RL Gene 67:105-115(1988).
[3] SEQUENCE FROM N.A.
RP MEDLINE=89025644; PubMed=3178743;
RX Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RL of human type I procollagen.";
RN Biochem. J. 253:919-922(1988).
[4] SEQUENCE FROM N.A.
RP MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
RX Maatta A., Bornstein P., Penttinen R.P.;
RA "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";
RL FEBS Lett. 279:9-13(1991).
[5] SEQUENCE FROM N.A.
RP MEDLINE=92157916; PubMed=1787829;
RX Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA Olsen A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
RL the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RN Matrix 11:375-379(1991).
[6] SEQUENCE FROM N.A.
RP MEDLINE=98107942; PubMed=9443882;
RX Korkko J., Ala-Kokko L., De Paape A., Nuytinck L., Earley J.,
RA Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
RT scanning by conformation-sensitive gel electrophoresis identifies only
RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I;
RL identification of common sequences of null-allele mutations.";
RN Am. J. Hum. Genet. 62:98-110(1998).
[7] SEQUENCE FROM N.A.
RP MEDLINE=98107942; PubMed=9443882;
RX Korkko J.M., Earley J.J., Nuytinck L., DePaape A., Prockop D.J.,
RA Ala-Kokko L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008885; Fib_collagen_C.
DR InterPro; IPR009041; FMP_SCCI.
DR InterPro; IPR010077; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; WWC; 1.
DR PROSITE; PS01208; WWC_1; 1.
DR PROSITE; PS0184; WWC_2; 1.
KW Collagen.
SQ SEQUENCE 1461 AA; 138629 MW; 9ACF6DE30EA78E21 CRC64;

```

Query Match          64.0%; Score 2233.5; DB 2; Length 1461;
Best Local Similarity 60.0%; Pred. No. 5.2e-81;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGRGGRGSRGPGADGVAGPKGAGERSGPPA 48
D 455 GPPGAGEGKRGARGEPGPTGLPGRGGRGSRGPGADGVAGPKGAGERSGPPA 514
QY 49 GPKGSPGAGRPGAGLPGAKGLTSGPSGPGDGTTPPGAGDGRPPGPPGARGQA 108
D 515 GPKGSPGAGRPGAGLPGAKGLTSGPSGPGDGTTPPGAGDGRPPGPPGARGQA 574
QY 109 GVMGFPKGKGAAGRPCKAGRGVPPGCAVGPAGKGEAGAQGPPGAGRGSGPPA 168
D 575 GVMGFPKGKGAAGRPCKAGRGVPPGCAVGPAGKGEAGAQGPPGAGRGSGPPA 634
QY 169 GSPFGQGLPGPAGPPGAGKGEQVPGDLGAPGSPGAGE-----PGP----- 212
D 635 GSPFGQGLPGPAGPPGAGKGEQVPGDLGAPGSPGAGE-----PGP----- 694
QY 213 -----TGLPGRGGRGSGRGGPGADGVAGPKGAGERSGPPGKSP 258
D 695 GAGAPGNDGAKGADGAGAPGAGSGQAPGQAPGQAPGAGLPGKDRGDAGPKGADGSP 754
QY 259 GE-----AGRPGAGLPGAK-----CLTSGPSGPGDGTTPPGAGDGRPP 300
D 755 GKQGVRLGTPGIPGPPGAGAPGDKGESGSPGAGTGAAGAPGDRGEPGPPGAGPAP 814
QY 301 GPPGPPGARGQ-----AGVMGFPKGKGAAGRPCKAGRGVPPGPPGA----- 341
D 815 GADQPKAGPEPDAGAKGADGPPGAPGAPGPGPIGNVAPGAKGARSAGPPGATGPP 874
QY 342 -----VGPAGKGEAGAQGPPGAGP---AGRGEGGAGSGPFGQLPGPAGPPGAGKP 393
D 875 GAAGRVGPPGSGNAGPPGPPGAGKEGKGRGPGTGPAGRGVPPGPPGAGKSGP 934
QY 394 GEQGVFDLGAAPGSPGAPGPGTGLPGRGGRGSGRGGPGADGVAGPKGAGERSP 453
D 935 GADGAPAGTPTGPGTLAGRGVVLPGQKRGGRGFFGLPGSPGEPGKQSGASGERGPP 994
QY 454 GPAGP---KSGPAGRPGAGLPGAKGLTSGPSGPGDGTTPPGAGDGRPPGPPGPP 510
D 995 GPMGPPGLAGPPGSGREGAPGAGSGRDRGSGPAGKDRGRTGPPGAPGAPGAPGV 1054
QY 511 GARGOAGVMGFPKGKGAAGRPCKAGRGVPPGCAVGPAGKGE-----AGAQ 558
D 1055 GPAGKSGDRGTGTPAGPAGVPGVARGPAGPQGRGKGTGEQGDRIKGRGFSGLQ 1114
QY 559 GPPGP-----AGPAGRGQGPPAGSP---GFGQLPGPAGPPGAGKPGEQGVP 603
D 1115 GPPGPPGSPGEGSGASGAPGPPGSGAGAPGKDLNGLPQIPGPPGRTGDAGVP 1174
QY 604 GDLAGPSPGAPG 616
D 1175 GPPGPPGPPGPPG 1187

RESULT 3
Q8N473
ID Q8N473 PRELIMINARY; PRT; 1464 AA.
AC
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen, preproprotein.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;

```

```

MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
Whiting R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
Krzyszinski M.L., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
Jones S.J., Marra M.A.;
"Generation and initial analysis of more than 15,000 full-length human
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC036531; AAH36531.1; -
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clq helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR009041; PMP_SCCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clq helix; 3.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS0184; VWF_C_2; 1.
DR Collagen.
SQ SEQUENCE 1464 AA; 139010 MW; B0581F8D1C89DDE8 CRC64;

Query Match          64.0%; Score 2233.5; DB 2; Length 1464;
Best Local Similarity 60.0%; Pred. No. 5.2e-81;
Matches 440; Conservative 23; Mismatches 153; Indels 117; Gaps 12;

QY 1 GPP-----GEPGPTGLPGRGGRGSRGPGADGVAGPKGAGERSGPPA 48
D 458 GPPGAGEGKRGARGEPGPTGLPGRGGRGSRGPGADGVAGPKGAGERSGPPA 517
QY 49 GPKGSPGAGRPGAGLPGAKGLTSGPSGPGDGTTPPGAGDGRPPGPPGARGQA 108
D 518 GPKGSPGAGRPGAGLPGAKGLTSGPSGPGDGTTPPGAGDGRPPGPPGARGQA 577
QY 109 GVMGFPKGKGAAGRPCKAGRGVPPGCAVGPAGKGEAGAQGPPGAGRGSGPPA 168
D 578 GVMGFPKGKGAAGRPCKAGRGVPPGCAVGPAGKGEAGAQGPPGAGRGSGPPA 637
QY 169 GSPFGQGLPGPAGPPGAGKGEQVPGDLGAPGSPGAGE-----PGP----- 212
D 638 GSPFGQGLPGPAGPPGAGKGEQVPGDLGAPGSPGAGE-----PGP----- 697
QY 213 -----TGLPGRGGRGSGRGGPGADGVAGPKGAGERSGPPGKSP 258
D 698 GAGAPGNDGAKGADGAGAPGAGSGQAPGQAPGAGLPGKDRGDAGPKGADGSP 757
QY 259 GE-----AGRPGAGLPGAK-----CLTSGPSGPGDGTTPPGAGDGRPP 300

```


Db 758 GKGVRLTGLTPIPPGAPAGDKGESGSPGAGTARGAPGDRGEPGPPGAPGAP 817
 Qy 301 GPPGPPGARGO-----AGVMGFPKGAAGBPAGKAGRGVPPGGA----- 341
 Db 818 GADGQPKAGBFGDAGKGDAGPAGPAGPPIGNVGAAPGAKGARGSGAPGATGFP 877
 Qy 342 -----VGPAGKDGAGAGGPPGAPG---AGERGQGPAGSPGFGGLPGAPGPPGAGK 393
 Db 878 GAARVVPDPSGNAGPPGPPGAPGKGGKGRGTGAPGPPGPPGPPGAPGK 937
 Qy 394 GEOQVPGDLGAPGSPGAGEPPTGLPDPGPPGPPGPPGPPGPPGPPGPPGPPG 453
 Db 938 GADGAPAGTPTGQIAGRGVGLPQRCRGFPGLPGSPGPPGPPGPPGPPGPP 997
 Qy 454 GPAGP-----KGSFGAGRPGAGLPGAKGLTSGPSGPPGDKTGPAGQDRGPPGPP 510
 Db 998 GPMGPPGLAGPSPGSGREGAPGAGSPGRCDSGPAKGDRTGTPAGPPGAPGAPG 1057
 Qy 511 GARGOAGVMGPPGPKGAGRPKAGRGVPPGPPGAPGAGKDG-----AGAQ 558
 Db 1058 GPAGSGDRGTGAPGAPGPPGPPGAPGAPGPPGPPGPPGPPGPPGPPGPP 1117
 Qy 559 GPPGP-----AGPARGERGQGPAGSP---GFQGLPGPAGPPGPPGPPGPPG 603
 Db 1118 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1177
 Qy 604 GDLGAPGPPGPPG 616
 Db 1178 GPPGPPGPPGPPG 1190

RESULT 4

CALL CANFA STANDARD; PRT: 1460 AA.
 AC Q9XSJ7;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE OCT-2004 (Rel. 45, Last annotation update)
 RT heterozygous alpha 1(I) chain precursor.
 GN Name=COL1A1;
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT OI ALA-208.
 RC TISSUE=Skin;
 RX MEDLINE=2102337; PubMed=11147834; DOI=10.1006/abbi.2000.2099;
 RA Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;
 RT "Sequence of normal canine COL1A1 cDNA and identification of a
 heterozygous alpha 1(I) collagen Gly208Ala mutation in a severe case of
 canine osteogenesis imperfecta."
 RL Arch. Biochem. Biophys. 384:37-46 (2000).
 CC -1- FUNCTION: Type I collagen is a member of group I collagen
 (fibrillar forming collagen).
 CC -1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
 CC -1- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -1- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta
 (OI).
 CC -1- SIMILARITY: Belongs to the fibrillar collagen family.
 CC -1- SIMILARITY: Contains 1 WFCC domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF153062; AAD34619.1; -;
 DR InterPro; IPR008160; Cig_helix.

DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR008885; Fib_collagen_C.
 DR InterPro; IPR001007; WFV_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; Cig_helix; 2.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR PROSITE; PS01208; WFV_C_1; 1.
 DR PROSITE; PS01208; WFV_C_2; 1.
 KW Collagen; Disease mutation; Extracellular matrix; Glycoprotein;
 KW Hydroxylation; Pyrrolidone carboxylic acid; Repeat; Signal;
 KW Structural protein.
 FT SIGNAL 1 22 By similarity.
 FT PROPEP 23 157 N-terminal propeptide.
 FT CHAIN 158 1214 Collagen alpha 1(I) chain.
 FT PROPEP 1215 1460 C-terminal propeptide.
 FT DOMAIN 34 92 WFV_C.
 FT DOMAIN 158 174 Nonhelical region (N-terminal).
 FT DOMAIN 175 1188 Triple-helical region.
 FT DOMAIN 1189 1214 Nonhelical region (C-terminal).
 FT SITE 741 743 Cell attachment site (Potential).
 FT SITE 1089 1091 Cell attachment site (Potential).
 FT MOD_RES 158 158 Pyrrolidone carboxylic acid (By similarity).
 FT MOD_RES 166 166 Allylsine (By similarity).
 FT MOD_RES 261 261 5-hydroxylysine (By similarity).
 FT MOD_RES 1160 1160 3-hydroxyproline (By similarity).
 FT CARBOHYD 261 261 O-linked (Gal..) (By similarity).
 FT CARBOHYD 1361 1361 N-linked (GlcNAc..) (By similarity).
 FT VARIANT 208 208 G -> A (in OI; severe).
 SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;
 Query Match 64.0%; Score 2231.5; DB 1; Length 1460;
 Best Local Similarity 60.2%; Pred. No. 6.2e-81;
 Matches 441; Conservative 20; Mismatches 155; Indels 117; Gaps 12;
 Qy 1 GPP-----GEPGPTGLPDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 48
 Db 454 GPPGPAEEGKRGARGEPGPTGLPDPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 513
 Qy 49 GPKGSPGEAGRPGRAGLPGAKGLTSGSPGSPGDKTGTGPPGAGQDGRGPPGPPG 108
 Db 514 GPKGSPGEAGRPGRAGLPGAKGLTSGSPGSPGDKTGTGPPGAGQDGRGPPGPPG 573
 Qy 109 GVMGFPKGAAGEPGKAGRGVPPGPAVGPAGKDGEGAGQGPAGPAGRGEGQGA 168
 Db 574 GVMGFPKGAAGEPGKAGRGVPPGPAVGPAGKDGEGAGQGPAGPAGRGEGQGA 633
 Qy 169 GSPGFQGLPGPAGPPGAGKRGEGVPGDILGAPGSPGAGE-----GPP----- 212
 Db 634 GSPGFQGLPGPAGPPGAGKRGEGVPGDILGAPGSPGARGERGPPGPPGPPGPP 693
 Qy 213 -----TGLPDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 258
 Db 694 GAGAPNDGAKGDAGAPGAPGSPGAGFOLQMPGEGAGLPGPKGDKGADGADGSP 753
 Qy 259 GE-----AGRPGEAGLPGAK-----GLTSGSPGSPGDKTGTGPPGAGQDGR 300
 Db 754 GKQGVRLTGLTPIPPGAPAGDKGAGSPGAPGTGARGAPGDRGPPGPPGPPGPP 813
 Qy 301 GPPGPPGARGO-----AGVMGFPKGAAGBPAGKAGRGVPPGGA----- 341
 Db 814 GADGQPKAGBFGDAGKGDAGPAGPAGPPIGNVGAAPGAKGARGSGAPGATGFP 873
 Qy 342 -----VGPAGKDGAGAGGPPGAPG---AGERGQGPAGSPGFGGLPGAPGPPG 393
 Db 878 GAARVVPDPSGNAGPPGPPGAPGKGGKGRGTGAPGPPGPPGPPGAPGK 933
 Qy 394 GEOQVPGDLGAPGSPGAGEPPTGLPDPGPPGPPGPPGPPGPPGPPGPPGPPG 453
 Db 938 GADGAPAGTPTGQIAGRGVGLPQRCRGFPGLPGSPGPPGPPGPPGPPGPP 993
 Qy 454 GPAGP-----KGSFGAGRPGAGLPGAKGLTSGPSGPPGDKTGPAGQDRGPPGPP 510

Db	299	GOMGRGLUGERGRPGAPGAPAGANGDAGTAAAGPPGPTGTPAGPPGPPGAVGAKGEAGPQ	358
QY	151	GPDPAGPAGERGEOCP-----AGSPGFQGLPGPA	180
Db	359	GRSGEGQGVGRGPPGPPAGAAGPAGNPGADGQPGAKGANGAPGAPGAPFPFGARGFS	418
QY	181	GPPGAGHPGQGVPGDILGAP-----GPSGA-----GBPGPT	213
Db	419	GPQGFPGGPPGKNSGEPCAPGSKGDTCAKGEPGVPVQVQPPGAGEGCKRGARSGPPT	478
QY	214	GLPDPGRGRGPPGRFRPGADVAGPKPACGERGSPGAPGKSGPGEAGRPGAEAGLPGAK	273
Db	479	GLPDPGRGRGPPGRFRPGADVAGPKPACGERGSPGAPGKSGPGEAGRPGAEAGLPGAK	538
QY	274	GLTSPGSPGPDGKTGTPPGAGQDGRPPGPPGARGQAGVGMFPQPKGAAGEPGKAGER	333
Db	539	GLTSPGSPGPDGKTGTPPGAGQDGRPPGPPGARGQAGVGMFPQPKGAAGEPGKAGER	598
QY	334	GVPGRGAVGPKDGEAGACGPPGPPGACGERGHPGAGSPGFGPLGPPAGPPGEGAGKP	393
Db	599	GVPGRGAVGPKDGEAGACGPPGPPGACGERGHPGAGSPGFGPLGPPAGPPGEGAGKP	658
QY	394	GEQVPGDILGAPGSPGAGE-----PGS-----TGLPGPP	423
Db	659	GEQVPGDILGAPGSPGAGERGFPGERGVQVQPPGAPRGANGAPCNDGAKDGAAGAP	718
QY	424	GERGPGSRPPGADGVAGPKPACGERGSPGAPGKSGPGE-----AGRPGAEGLP	474
Db	719	GSQAPGLQGMPPGRGAAGLFGPKGDRDAGPKGADGSPGKDGVRGLTGPIGPPGAPAG	778
QY	475	GAK-----GLTSPGSPGPDGKTGTPPGAGQDGRPPGPPGARGQ-----	515
Db	779	GDKGESGSPGAPGTGARGAPDRGRGPPGPPGAPGAPGADGQPKAGEPCDGAAGKDA	838
QY	516	-----AGVMGPPGPKAAGSPKAGERGVPPGGA-----VGPAGKDGCAAGAGPP	561
Db	839	GPPGAPGAPGPPGPIGNVGAAPKAGARGSPGATGPPGAAGRVGPPGPSNAGPPGPP	898
QY	562	GPAGP---AGERGHPGAGSPGFGPLGPPAGPPEAGKPGGQVPGDLGAPGSPGAG	616
Db	899	GPAGKEGKGRGETGTGAPRGVEVGPVPPGPPGAGEKSPGADGAPAGFTGTPGQGIAG	956
RESULT 6			
Q63079	PRELIMINARY; PRT; 1453 AA.		
ID	Q63079		
AC	Q63079;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	Collagen alpha1 (Fragment).		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxID=10116;		
EN	[1]		
RP	SEQUENCE FROM N. A.		
RC	STRAIN=Sprague-Dawley; TISSUE=Bone, and Tooth;		
RX	MEDLINE=99163824; PubMed=10065941;		
DR	GO; GO:0005581; C:collagen; IEA.		
DR	GO; GO:0005737; C:cytoplasm; IEA.		
DR	GO; GO:0005201; F:extracellular matrix structural constituent; IEA.		
DR	GO; GO:000617; P:phosphate transport; IEA.		
DR	InterPro; IPR008161; C1g_nelix.		
DR	InterPro; IPR008160; Collagen.		
DR	InterPro; IPR000885; Fib collagen_C.		
DR	InterPro; IPR009041; PMP SGC1.		
DR	InterPro; IPR001007; VWF_C.		

RESULT 7

CALL MOUSE STANDARD; PRT; 1453 AA.

ID CALL MOUSE STANDARD; PRT; 1453 AA.

AC P11087; Q60635;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 25-OCT-2004 (Rel. 45, Last annotation update)

DE Collagen alpha 1(I) chain precursor.

DE Names=Colla1; Synonyms=Colal;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_Taxid=10090;

[1]

RN SEQUENCE FROM N.A.

RP STRAIN=FVB/N;

RC MEDLINE=96033240; PubMed=8535610;

RX Li S.W., Killian J., Prockop D.J.;

RA "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain

RT of type I procollagen.";

RL Matrix Biol. 14:593-595(1995).

[2]

RN SEQUENCE OF 518-1128 FROM N.A.

RP MEDLINE=86137403; PubMed=3841523; DOI=10.1016/0378-1119(85)90329-4;

RX French B.T., Lee W.-H., Maul G.G.;

RA "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I) collagen

RT protein.";

RL Gene 39:311-312(1985).

[3]

RN SEQUENCE OF 735-1130 FROM N.A.

RP MEDLINE=83141374; PubMed=6298597;

RX Monson J.M., Friedman J., McCarthy B.J.;

RA "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:

RT evidence for a mouse B1 element within the gene.";

RL Mol. Cell. Biol. 2:1362-1371(1982).

[4]

RN SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.

RP MEDLINE=83157109; PubMed=6219867;

RX Monson J.M., McCarthy B.J.;

RA "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:

RT evidence for insertions or deletions in gene coding sequences.";

RL DNA 1:59-69(1981).

[5]

RN SEQUENCE OF 1442-1453 FROM N.A.

RP MEDLINE=88124276; PubMed=3340560;

RX Mooslehner K., Harbers K.;

RA "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size of

RT the 3'-untranslated region.";

RL Nucleic Acids Res. 16:773-773(1988).

CC -1- FUNCTION: Type I collagen is a member of group I collagen

CC (fibrillar forming collagen).

CC -1- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.

CC -1- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and

CC bones. In bones the fibrils are mineralized with calcium

CC hydroxyapatite.

CC -1- PTM: Prolines at the third position of the tripeptide repeating

CC unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -1- SIMILARITY: Belongs to the fibrillar collagen family.

CC -1- SIMILARITY: Contains 1 WFCC domain.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on

CC use by non-profit institutions as long as its content is in no

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements>

CC or send an email to license@isb-sib.ch).

CC

DR ENBL; U08020; AAA88912.1; -

DR ENBL; X15896; CAA33904.1; -

DR ENBL; M14423; AAA37333.1; -

DR ENBL; M17491; AAA37334.1; -

DR ENBL; X06753; CAA29927.1; -

DR EMBL; K03036; AAA37332.1; --
DR EMBL; K03029; AAA37332.1; JOINED.
DR EMBL; K03030; AAA37332.1; JOINED.
DR EMBL; K03031; AAA37332.1; JOINED.
DR EMBL; K03032; AAA37332.1; JOINED.
DR EMBL; K03033; AAA37332.1; JOINED.
DR EMBL; K03034; AAA37332.1; JOINED.
DR EMBL; K03035; AAA37332.1; JOINED.
DR PIR; S57243; S21626.
DR MGD; MGI:88467; Coll1a1.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008180; Collagen.
DR InterPro; IPR00885; Fib collagen_C.
DR InterPro; IPR009041; PMP SGC1.
DR InterPro; IPR001007; WVF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 3.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; WVC; 1.
DR PROSITE; PS01208; WVF_C; 1.
DR PROSITE; PS0184; WVF_C; 1.
DR Collagen; Extracellular matrix; Glycoprotein; Hydroxylation;
DR Pyrrolidone carboxylic acid; Repeat; Signal; Structural protein.
FT SIGNAL 1 22
FT PROPEP 23 151 N-terminal propeptide.
FT CHAIN 152 1207 Collagen alpha 1(I) chain.
FT PROPEP 1208 1453 C-terminal propeptide.
FT DOMAIN 29 87 WVF_C
FT DOMAIN 152 167 Nonhelical region (N-terminal).
FT DOMAIN 168 1181 Triple-helical region.
FT DOMAIN 1182 1207 Nonhelical region (C-terminal).
FT MOD_RES 152 152 Pyrrolidone carboxylic acid (By
FT similarity).
FT MOD_RES 160 160 Alanine (By similarity).
FT MOD_RES 234 234 5-hydroxylysine (By similarity).
FT MOD_RES 1153 1153 3-hydroxyproline (By similarity).
FT CARBOHYD 56 56 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 254 254 O-linked (Gal...) (By similarity).
FT CARBOHYD 1354 1354 N-linked (GlcNAc...) (By similarity).
FT SITE 734 736 Cell attachment site (Potential).
FT SITE 1082 1084 Cell attachment site (Potential).
FT CONFLICT 1450 1450 A -> V (in Ref. 5).
FT SEQUENCE 1453 AA; 137944 MW; 3B802E535DF81808 CRC64;
Query Match 63.2%; Score 2205.5; DB 1; Length 1453;
Best Local Similarity 58.9%; Pred. No. 6.6e-80;
Matches 432; Conservative 22; Mismatches 162; Indels 117; Gaps 11;
QY 1 GPP-----GEPGTLGPPGGERGSGRGGFPGADGAGKPGAGRGSPGPA 48
Db 447 GPPGAGEGKRGAGEGPGSLGPPGGERGSGRGGFPGADGAGKPGAGRGSPGPA 506
QY 49 GPKSGPGEAGRGAGLPGAKGLTSGSPGPDGKTGTPGAGQGRGPPGPPGARGQA 108
Db 507 GPKSGPGEAGRGAGLPGAKGLTSGSPGPDGKTGTPGAGQGRGPPGPPGARGQA 566
QY 109 GVMGPPGKGAAGEGPKAGRGVPGGAVGAGKDGAGAGQPPGAPGAGEGCGPA 168
Db 567 GVMGPPGKGAAGEGPKAGRGVPGGAVGAGKDGAGAGQPPGAPGAGEGCGPA 626
QY 169 GSPFGQLGPPGAPGAGEGPKAGRGVPGGAVGAGKDGAGAGQPPGAPGAGEGCGPA 212
Db 627 GSPFGQLGPPGAPGAGEGPKAGRGVPGGAVGAGKDGAGAGQPPGAPGAGEGCGPA 686
QY 213 -----TCLPGPPGERGSGRGGFPGADGAGKPGAGRGSPGPA 258
Db 687 GNNGAPGNDGAGKDGTPGAGPAGKGLTSGSPGPDGKTGTPGAGQGRGPPGPPGARGQA 746
QY 259 GE-----AGRPGEAGLPGAKGLTSGSPGPDGKTGTPGAGQGRGPPGPPGARGQA 300
Db 747 GKDAGRLTGPIGPPGAPGAFKDGAGRGVPGGAVGAGKDGAGAGQPPGAPGAGEGCGPA 806

QY 301 GPPGPPGARGQ-----AGVMGPPGKGAAGEGPKAGRGVPGGPA----- 341
Db 807 GADGQPGAKGEPGDTGVKGDAGPPGAPGAPGPPGIGNVGAPGKPGAGAPPGATGPP 866
QY 342 -----VGPAGKDGAGAGQPPGAPG-----AGERGQGPAGSPGFGGLPGAPGPPGEGAKP 393
Db 867 GAAGRVGPPGSGNAGPPGPPGVCKGKGKPRGETGAPGPPGPPGPPGPPGPPGPPGPPG 926
QY 394 GEQGVGDLGAPGSPGAPGEPGTLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 453
Db 927 GADGAPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 986
QY 454 GPAGP-----KGSPPGAGRPGEAGLPGAKGLTSGSPGPDGKTGTPGAGQGRGPPGPPG 510
Db 987 GPMGPPGLAGPPGSGREGSPGAEGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1046
QY 511 GARGQAGVMGPPGKGAAGEGPKAGRGVPGP-----PGAV 546
Db 1047 GPAGKNGDRGETGPPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 1106
QY 547 GPAGKDGAGAGQPPGPPGAPGAGEGCGPAGSP-----GQGLPGPAGPAGPAGPAGPAG 603
Db 1107 GPPSGPSGQGGSGAGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1166
QY 604 GDLGAPGSPGAPG 616
Db 1167 GPPGPPGPPGPPG 1179
PRELIMINARY; PRT; 1453 AA.
RESULT 8
Q810J9
ID Q810J9 PRELIMINARY; PRT; 1453 AA.
AC Q810J9
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
GN Name=Coll1a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC STRAIN=FROM N.A.
EX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting R., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RC STRAIN=FROM N.A.
EX SEQUENCE=Colon;
RA Strausberg R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC050014; AAH50014.1; --

DR MGI:88467; Colial.
DR GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib_collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00214; VWC; 1.
DR SMART; SM00038; COLFI; 1.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS0184; VWF_C; 2; 1.
DR Collagen.
KW Collagen.
SQ SEQUENCE 1453 AA; 138032 MW; 0B7F06BBB9A1D5EA CRC64;

Query Match 63.2%; Score 2205.5; DB 2; Length 1453;
Best Local Similarity 58.9%; Pred. No. 6.6e-80;
Matches 432; Conservative 22; Mismatches 162; Indels 117; Gaps 11;

Qy 1 GPP-----GEPCTGLPGPGERGSGRSGFPAGDGVAGPKPAGERGSPGA 48
Db 447 GPPPAGEEGKRGERGEPPGSLFPGERGSGRSGFPAGDGVAGPKPAGERGATGPA 506

Qy 49 GPKSGPGEAGRPGAGLPGAKGLTGSPPGPDGKTGTPPGAGDGRPPGPPGPGARQQA 108
Db 507 GPKSGPGEAGRPGAGLPGAKGLTGSPPGPDGKTGTPPGAGDGRPPGPPGPGARQQA 566

Qy 109 GVMFPFGKGAAGPCKAGRGVPGPCAVGAPGAKDGEAGAGPPGAPGAGERGEGQPA 168
Db 567 GVMFPFGKGAAGPCKAGRGVPGPCAVGAPGAKDGEAGAGPPGAPGAGERGEGQPA 626

Qy 169 GSPFGQLPGPAGPPGKAGRGVPGDLCAPGSPGAG------RGP---- 212
Db 627 GSPFGQLPGPAGPPGKAGRGVPGDLCAPGSPGAGRGVPGGVPGRGVPAGPR 686

Qy 213 -----TGLPFGPGERGSGRSGFPAGDGVAGPKPAGERGSGPPGPKGSP 258
Db 687 GNNGAPGNDGAKGDTGAPGFSQAGLCQMPGERGAAGLPKGDGADGPKGADGSP 746

Qy 259 GE-----AGRPGAGLPGAKGLTGSPPGPD-----GKTGPPGPPAGQDGRP 300
Db 747 GKDAGLGTGTPGPPGAPAGDGEAGSPGPGTCTGARGAPGDRGEAGPPGPPGAGFAGPP 806

Qy 301 GPPGPPGARGO-----AGVMGFPKPKGAAGPCKAGRGVPGPPGA---- 341
Db 807 GADQPGKAGPDTGVKGDAGPPGAPGPPGPIGNVAGPPKPKGAAGPPGATGFP 866

Qy 342 -----VGPAGKDGAGAGQPPGAPG---AGERGEQPPAGSFGQLPGPAGPPGAGPK 393
Db 867 GAAGRVPPGPPGSGNAGPPGPPGVGKGGKGRGTGTPAGRPGEVPPGPPGAGERGSP 926

Qy 394 GEQVGDGLAPGSPGAPGPTGLPGPPGERGSGRSGFPAGDGVAGPKPAGERGSP 453
Db 927 GADGPPAGSPGTPGQIAGQGVVGLPGQERGERFGLPGPSGPFQKQPSGSSGERGPP 986

Qy 454 GPAGP---KGSPPGAGRGEAGLPGAKGLTGSPPGPDGKTGTPPGAGQDGRGPPGPP 510
Db 987 GPMGPPGLAGPPGSGRSGSPGAGSGRDRDAPGAKGDRGTGAPGPPGAPGAPGPPV 1046

Qy 511 GARGQAGVMGFPKPKGAAGPCKAGRGVPGP-----PGAV 546
Db 1047 GPAGKNGDRGTGTPAGPAGPIGPAGARCAPGPPGPRGDKGTGQDRIKHKRFGSLQ 1106

Qy 547 GPAGKDGAGAGQPPGPPGAPGAGERGEGQPPAGSP---GFQGLPGPAGPPGAGKPCGEGVP 603
Db 1107 GPPSGPSGPPGQSGASGAPGPPGSPGSGAGSGPGKDLGLNGLPGLPIGPPGPRGRTGDSGA 1166

Qy 604 GDLGAPGSPGAG 616
| | | | |

Db 1167 GPPGPPGPPGPPG 1179

RESULT 9
Q6PCL3 PRELIMINARY; PRT; 1225 AA.
AC Q6PCL3;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Colial protein.
GN Name=Colial;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
EX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Basak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
RA Strausberg R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib_collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 13.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS0184; VWF_C; 2; 1.
KW Collagen.
SQ SEQUENCE 1225 AA; 117860 MW; B6B86CB4457F4D9 CRC64;

Query Match 62.4%; Score 2176.5; DB 2; Length 1225;
Best Local Similarity 56.7%; Pred. No. 8.2e-79;
Matches 431; Conservative 28; Mismatches 157; Indels 144; Gaps 13;

Qy 1 GPPGEPGPTGLPDPGPPG-----RGGPGSRGFPAGDGVAGPKPAGERGSPGA 48
| | | | |
Db 180 GPPCAPCPQCFQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPQ 239
| | | | |
Qy 49 GPKSGPGEAGR-----GEAGLPGAKGLTGSPPGPPGPPGPPGPPGPPGPPGPPG 93
| | | | |

DR SMART: SM00214; VWC: 1.
DR PROSITE; PS01208; VWC_1; 1.
DR PROSITE; PS50184; VWC_2; 1.
KW Collagen; Direct protein sequencing; Extracellular matrix;
KW Glycoprotein; Hydroxylation; Pyrrolidone carboxylic acid; Repeat;
KW Signal; Structural protein.
FT SIGNAL 1 22
FT PROPEP 23 151 N-terminal propeptide.
FT CHAIN 152 1205 Collagen alpha 1(i) chain.
FT PROPEP 1206 1453 C-terminal propeptide.
FT DOMAIN 31 89 VWC.
FT MOD_RES 152 152 Pyrrolidone carboxylic acid.
FT MOD_RES 160 160 Alllysine (By similarity).
FT MOD_RES 254 254 5-hydroxylysine (By similarity).
FT MOD_RES 851 851 5-hydroxylysine (Potential).
FT MOD_RES 1081 1081 Hydroxyproline (Potential).
FT MOD_RES 1097 1097 5-hydroxylysine (Potential).
FT MOD_RES 1153 1153 3-hydroxyproline.
FT CARBOHYD 254 254 O-linked (Gal...) (By similarity).
FT CARBOHYD 1354 1354 N-linked (GlcNAc...) (By similarity).
FT CONFLICT 1187 1187 F -> L (in Ref. 5).
FT CONFLICT 1441 1441 Q -> H (in Ref. 6).
SQ SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;

Query Match 61.7%; Score 2151.5; DB 1; Length 1453;
Best Local Similarity 57.8%; Pred. No. 8.8e-78;
Matches 424; Conservative 17; Mismatches 175; Indels 117; Gaps 10;

QY 1 GPP-----GERPGTGLPPGCGRGGSPGAGDGVAGPKGAGRGSPGPA 48
DB 447 GPPGAGEGKRGAGEGPGAGLPGPAGERGAPSRGPGGADGAGPKGPGGSGPAG 506

QY 49 GPKGSPGAGRGGAGLPGAKGLTSGSPGPGDGTGTPGAGDGRGPPGPPGARGQA 108
DB 507 GPKGSPGAGRGGAGLPGAKGLTSGSPGPGDGTGTPGAGDGRGPPGPPGARGQA 566

QY 109 GVMGPPGKGAAGRGGAGLPGAGVAGPKGAGDGAAGGPPGAGRGGEQGA 168
DB 567 GVMGPPGKGAAGRGGAGLPGAGVAGPKGAGDGAAGGPPGAGRGGEQGA 626

QY 169 GSPGFQGLPGAGPGEAGKPGDGLGAGPSPGAGE-----PQP----- 212
DB 627 GAGPFGQLPGAGPGEAGKPGDGLGAGPSPGAGE-----PQP----- 686

QY 213 -----TGLPGPGEGRGGSPGPGAGDVAGPKGAGRGSGP----- 251
DB 687 GAGAPGNDGAKGADAGAPGAGNEGPPGLEMGPERGAAGLPGAKGDRGDFGPKGADGAP 746

QY 252 -----AGKPSGEGAGRP-----GEAGLPGAKGLTSGSPGPDGKTGTPGPA----- 294
DB 747 GKDGRLGLTPIGPPGAGAPGDKGEAGPPGAGTGAAGAPGDRGEPGPPGAGFAGPP 806

QY 295 -----GQDRGPPGPPGARGGAGVGMGPPGKGAAGRGGKAGRGVPP 336
DB 807 GADQPGAKGTGAGAKGADGAPGAGTGAAGVAGVAGPAGPKGAGSGAPGATGFP 866

QY 337 GPPGAVGAGKDGAGAGAGPPGAGPAGE-----RGEQGPAGSPGFGQLPGAGPGEAGKP 393
DB 867 GAAGRVGPPGSGNIGLPGPPGAGKZSGKPRGETGAGRGPPGAGPPGPPGEGKSP 926

QY 394 GEQGVPGDLGAPGSPGAGEPGTGLPDPGERGGPSRGGFFGADGVAGPKGAGRGSP 453
DB 927 GADGPIGAGTTPGQIAGQRGVWGLPQQRGERGPPGLPGSPGEPKGQSPGASGERGPP 986

QY 454 GPAGP---KSGPGEAGRGEAGLPGAKGLTSGSPGPDGKTGTPGAGDGRGPPGPP 510
DB 987 GPMGPPGLAGPGEAGREGAPGAEAGRDAAGAPKGRGTGTPGAGPPGAPGAPGAPGV 1046

QY 511 GARGGAGVGMGPPGKGAAGRGGKAGRGVPPGAGVAGPAGKDG-----AGAQ 558
DB 1047 GPAGKNGRGTGTPGAGPAGPPGAGARGPAGPQGRGDKGTGEGDGRKMGKRGFSGLQ 1106

QY 559 GPP-----GPAGPAGERGEGQGPAGSPFGQLPGAPPPGAEAGKPGEGQGV 603

DB 1107 GPPGPPGAGGEGQSPGASGAGPAGRGPPGSGAGAKGDLNGLPGIPGPPGRGRTGEGVP 1166
QY 604 GDLGAPGSPGAPG 616
DB 1167 GPPGPPGPPGPPG 1179

RESULT 11
QY YIB4
ID QY YIB4 PRELIMINARY; PRT; 1450 AA.
AC QY YIB4
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen.
OS Cynops pyrrhogaster (Japanese common newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX NCBI_TaxID=8330;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Regenerate forelimbs;
RX MEDLINE=99407244; PubMed=10474166;
EX DOI=10.1002/(SICI)1097-0177(199909)216:1<59::AID-DVDY8>3.3.CO;2-2;
RA Ashihara K., Obara M., Yoshizato K.;
RT "Expression of genes of type I and type II collagen in the formation
and development of the blastema of regenerating newt limb.";
RL Dev. Dyn. 216:59-71(1999).
DR EMBL; AB015438; BAA36973.1; -;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR009041; PMP_SCCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD000007; Clg helix; 4.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWC_1; UNKNOWN_1.
DR PROSITE; PS50184; VWC_2; 1.
KW Collagen.
SQ SEQUENCE 1450 AA; 137563 MW; ABF8A74841B87B7C CRC64;

Query Match 58.6%; Score 2043.5; DB 2; Length 1450;
Best Local Similarity 54.7%; Pred. No. 1.6e-73;
Matches 401; Conservative 31; Mismatches 184; Indels 117; Gaps 11;

QY 1 GPP-----GEPGPTGLPGPGRGCGSPGSRGPPGADGVAGPKGAGRGSPGPA 48
DB 444 GPPGSGEGRKGRGSGPAGPPGPPGAGRGSGPSPGSDGASGPKGAPGERSVGA 503

QY 49 GPKGSPGAGRGGAGLPGAKGLTSGSPGPGDGTGTPGAGDGRGPPGPPGARGQA 108
DB 504 GPKGSGESRGGPPGPPGAGKGLTSGSPGPGDGTGTPAGAGDGHPPGPPGARGQS 563

QY 109 GVMGPPGKGAAGRGGAGRGVPPGAVGAPGAGKDGAGAGQSPGPPGAGRGGEQGA 168
DB 564 GVMGPPGKGAAGRGGAGRGVAGPAGTGAAGAGAGQSPGPPGSPGSGERGEQGA 623

QY 169 GSPGFQGLPGAGPGEAGKPGDGLGAPGSPGAGE-----PGPTGLPGP- 218
DB 624 GSPGFQGLPGSPGAGEAGKPGEQGAPGADGAGPPGSPGRGPPGRRGGQSPGAGQGR 683

QY 219 -----PGERGGPSRGFP----- 231
DB 684 GSPGSPNDGAKGAGAGAGAPGGRGPPGLQMPGERSAGMPGAKGDRGDAGTGDAGAP 743

KW	Collagen.
SQ	SEQUENCE 1445 AA; 137251 MW; F59BB550C9873F04 CRC64;
	Query Match 58.6%; Score 2042.5; DB 2; Length 1445;
	Best Local Similarity 49.7%; Pred.No. 1.7e-73;
	Matches 421; Conservative 34; Mismatches 161; Indels 231; Gaps 18;
QY	1 GPGEPTGLPGPPGERGGPS-----RGFFGADGVAGPKGA-----GER 42
Db	221 GKGRFGERFPQGARGLPGTAGLPGMKHGRFNGLDGAKGDTGPAGPKGEFGNGEN 280
QY	43 GSPAPAGKSGCEARPCBEAGLPGAKGLTSPDGSPPGPKTKTPPP-----A 90
Db	281 GAFQVVPRLPGERGRPGSPGAGARGNDGTPGAAGPFPTGTPGPPFGVGSPKGDA 340
QY	91 GDGRFPFPFGARQAQAVMGFPFGKAAGSPGKAGERGVPGPCGAVGPAKDGEGAAQ 150
Db	341 GPOSGRPGDPGARCEPGAPQAGPAGSAGNPGTDGQFGKATGATGAIAGAPFPGAR 400
QY	151 GPPGAPGAGEREQCPAGSPGFQGLPGPAGPPGIAKPGEGVPGDLGAPGPSGAPBP 210
Db	401 GAPPGPGGSGPFGKNNNEPQAQNKGBFAGKESGPASQCPGPPCEEKGRSRGP 460
QY	211 GTGLFPFPERGGPSGRFPFADGVAGPKPAGERGSPGPAGPKSGPCEAGRPEAGLP 270
Db	461 GPSGPPGAGERGAFGSRGFFPADAGAGFKGPPGERGVPVSGAPKSGPESGRPEGPL 520
QY	271 GAKGLTSGSPGPDGKTGPPGAPQODQRPPGPPGARGAQVGMFPKPKAAGEPGKA 330
Db	521 GAKGLTSGSPGPDGKTGPAGAPQODQRPPGPPGARGQSGVMGFPKGAAGEPGKP 580
QY	331 GERGVGPPGCAVGPAGKDGEAGAQQPPCAPGAPGAGERGGPAGSPGFOCLPGPAGPGEA 390
Db	581 GERGVAGFPACVAPCKDGEAGAQQPPCAPGAGERGGPAGSPGFOCLPGSPGAPGES 640
QY	391 GKPGSQGVPGD-----LGAPGSPG-----AGEP 414
Db	641 GKPGSQGAPGDVGPSPGASRGSRGFPGERGAI GPPGPGPRGANGAPCNDGAKGEAGAP 700
QY	415 -----GPTGLPGPPGERG-----GP-GSRGFPFADGV-----440
Db	701 GAPCGQSGSLQGWMPGERGAGGLPCA KD RD DQ GPKADGAPGDKVRGLTGRTGTPPGPG 760
QY	441 -----AGPKGA-----GERSGPPAGP-----KSGPGEAGRP 468
Db	761 GAPDKGEAGPAGPAGPTGSRGAPGERGEPGSPGAPGPPGADGQPAKGEQGDAGPK 820
QY	469 GEAGLPGAKGLTGP-----GSPGPDGKTGPP 495
Db	821 GDAGPPGAAGTGAFFGAVGATGPKGARGPAGPSTGPPGAAGRVGPPGSPGNAGPP 880
QY	496 -----GPAGQDRP-----GPPGP-----PGARGQAVMGFPKPGKAA 528
Db	881 GPSGAPGKEGQKPRGETGAGRCFEEGAAAGPPGPGSEKSGSDGAPAGPIPGOGIA 940
QY	529 GE-----PKVABRGVPPGAVGPAKDC-----AGAQPPIP-----563
Db	941 GTRTVGLPQGRGERGFGPLPGTTEPGKQSGSPSGERGPPGPGSPGLAGPPGPRE 1000
QY	564 -----AGPAGERGGPAGSPFOGLPGPAGPPGEAKHPGQGVPGDLGAP 609
Db	1001 GSPGSGSPGDSAGPKGRGSSGAPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 1060
QY	610 GPSGAPG 616
Db	1061 GPAGPAG 1067

RESULT 13
Q640B2
ID Q640B2 PRELIMINARY; PRT; 1449 AA.
AC Q640B2;
DT 25-OCT-2004 (TRENBLrel. 28, Created)

QY	340	GAVGPAKGGEAGACQAQCPGPAGPAGE-----RGEQGPGSPGFQGLPGAGPGRGEAKGP	393
Dd	864	GATGRVYGGPPAGNCGPPGSPGSPAGKEAGKGRGETGTGAPRGPEGGAAGPAGPPGEGKSP	923
QY	394	GEQGVDPDLCAGPSPGAPGEPGTGLPGPPGERGGPGSRGFPFGADGVAGPKGAPAGERGSP	453
Dd	924	GSDGPAGAPGCIIPGQCVAGSRSRTVGLPGMRGERGFSGLPGVPGEHQPQKQGPSFSGERGPP	983
QY	454	GPAGP---KGSPOEAGR---PGBAGLPGAAGKGLT-----GSPGSPGPDGKT	492
Dd	984	GPSGPPCLAGPPCEAGREGAPGSEGAPGRDGAAAGPKGDRCGGGAGPAGPGAPGAPGAFGPV	1043
QY	493	GPPGPAGQDGRGPPGPPGARGOAGVMGPPGKAAGEPGKAGE-----RGVPGPP	543
Dd	1044	GPAGKSGDRGETGAPGAPGAPGAGVAGARGPAGPQPRGDKGEAGEQBGRGMKHGRFNWPA	1103
QY	544	GAVGPAKGGEAGACQAQCPGPAGPAGERGEOGPAGSPGFQGLPGPAGPPGAGPKPGEQGV	603
Dd	1104	GPPGPPSSGEOGPSCASGAPAGPRGPPGSSGNPKDANGNLPGPIGPPGPRGRTGDVGPA	1163
QY	604	GDLAGAFPSPGPAG	616
Dd	1164	GPPGPPGPPGPPG	1176

RESULT 14

Q802B5	PRELIMINARY;	PRT; 1449 AA.
ID	Q802B5	
AC	Q802B5;	
DT	01-JUN-2003 (TEMBLrel. 24, Created)	
DT	01-JUN-2003 (TEMBLrel. 24, Last sequence update)	
DT	01-MAR-2004 (TEMBLrel. 26, Last annotation update)	
DE	Colial-prov protein.	
OS	Xenopus laevis (African clawed frog).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;	
OC	Xenopodinae; Xenopus.	
OX	NCBI_TaxID=8355;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Embryo;	
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;	
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,	
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,	
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,	
RA	Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,	
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,	
RA	Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,	
RA	Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,	
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,	
RA	Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,	
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,	
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,	
RA	Faney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,	
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,	
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,	
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buttefield Y.S.,	
RA	Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,	
RA	Jones S.J., Marra M.A.;	
RT	"Generation and initial analysis of more than 15,000 full-length human	
RT	and mouse cDNA sequences."	
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Embryo;	
RX	MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;	
RA	Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,	
RA	Richardson P.;	
RT	"Genetic and genomic tools for Xenopus research: The NIH Xenopus	
RT	initiative."	
RL	Dev. Dyn. 225:384-391(2002).	
RN	[3]	
RP	SEQUENCE FROM N.A.	

Qy	583	GLFGPAGPGGAGKPGCEQGVPCDLCAGPSPGAG	616
Db	1098	GFNGSPGPPPGSSGEOGSPGASGAPGPRGPPG	1131
RESULT 15			
ID	CALL	RAT	
AC	P02454	P02455;	PRT; 671 AA.
DT	21-JUL-1986	(Rel. 01, Created)	
DT	01-FEB-1994	(Rel. 28, Last sequence update)	
DT	25-OCT-2004	(Rel. 45, Last annotation update)	
DE	Collagen alpha 1(I) chain (Fragments).		
GN	Name=Colla1;		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
NCBI_TaxID	10116;		
RP	[1]		
RP	SEQUENCE OF 1-19.		
RA	MEDLINE=69155173; PubMed=5777344;		
RA	Bornstein P.;		
RT	"Comparative sequence studies of rat skin and tendon collagen. II. The		
RT	absence of a short sequence at the amino terminus of the skin alpha-1		
RT	chain.";		
RL	Biochemistry 8:63-71(1969).		
RP	[2]		
RP	SEQUENCE OF 5-19.		
RA	MEDLINE=67162268; PubMed=5337886;		
RA	Kang A.H., Bornstein P., Piez K.A.;		
RT	"The amino acid sequence of peptides from the cross-linking region of		
RT	rat skin collagen.";		
RL	Biochemistry 5:788-795(1967).		
RP	[3]		
RP	SEQUENCE OF 20-55.		
RA	MEDLINE=67165368; PubMed=4290711;		
RA	Bornstein P.;		
RT	"The incomplete hydroxylation of individual prolyl residues in		
RT	collagen.";		
RL	J. Biol. Chem. 242:2572-2574(1967).		
RP	[4]		
RP	SEQUENCE OF 56-102.		
RA	MEDLINE=71263178; PubMed=4327399;		
RA	Butler W.T., Ponds S.L.;		
RT	"Chemical studies on the cyanogen bromide peptides of rat skin		
RT	collagen. Amino acid sequence of alpha 1-CB4.";		
RL	Biochemistry 10:2076-2081(1971).		
RP	[5]		
RP	SEQUENCE OF 103-139.		
RA	MEDLINE=70085124; PubMed=5411206;		
RA	Butler W.T.;		
RT	"Chemical studies on the cyanogen bromide peptides of rat skin		
RT	collagen. The covalent structure of alpha 1-CB5, the major hexose-		
RL	Biochemistry 9:44-50(1970).		
RP	[6]		
RP	SEQUENCE OF 140-238.		
RA	MEDLINE=72136131; PubMed=4335087;		
RA	Balian G., Click E.M., Bornstein P.;		
RT	"Structure of rat skin collagen alpha 1-CB8. Amino acid sequence of		
RT	the hydroxylamine-produced fragment HA1.";		
RL	Biochemistry 10:4470-4478(1971).		
RP	[7]		
RP	SEQUENCE OF 239-418.		
RA	MEDLINE=73006942; PubMed=4342027;		
RA	Balian G., Click E.M., Hermodson M.A., Bornstein P.;		
RT	"Structure of rat skin collagen alpha 1-CBB. Amino acid sequence of		
RT	the hydroxyl amine-produced fragment HA2.";		
RL	Biochemistry 11:3798-3806(1972).		
RP	[8]		
RP	SEQUENCE OF 419-567.		
RA	MEDLINE=74271984; PubMed=4366532;		
RA	Butler W.T., Underwood S.P., Finch J.E. Jr.;		

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:15:59 ; Search time 134.148 Seconds
(without alignments)
2367.017 Million cell updates/sec

Title: US-10-658-989A-3

Perfect score: 4640

Sequence: 1 GPPGEPGTLGPPGGERGG.....GEQVPGDLGAPGSPGAGG 821

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_l6Dec04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003as.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4640	100.0	821	Adm48392	Human rec
2	3482	75.0	617	Adm48391	Human rec
3	3482	75.0	617	Adq26217	Human gel
4	2765.5	59.6	1449	Aae02535	Porcine a
5	2765	59.6	1057	Aay84544	A human c
6	2765	59.6	1058	Aay84403	Amino aci
7	2765	59.6	1107	Aay84540	Amino aci
8	2765	59.6	1171	Aay84538	A chimeri
9	2765	59.6	1464	Aaw68485	Human rec
10	2765	59.6	1464	Ad445059	Human pro
11	2765	59.6	1464	Ad445055	Human pro
12	2765	59.6	1464	Ad445051	Human pro
13	2765	59.6	1464	Adri6800	Human col
14	2765	59.6	1464	Adri6425	Human col
15	2765	59.6	1464	Adri99144	Collagen,
16	2761	59.5	1461	Abg93947	Human pol
17	2761	59.5	1464	Aau14136	Human nov
18	2759	59.5	1057	Aay84541	Amino aci
19	2757	59.4	1388	Aay84539	Amino aci
20	2755	59.4	1464	Aab82454	Human pro
21	2755	59.4	1464	Abb90764	Human tum
22	2755	59.4	1464	Abp68610	Human pan
23	2755	59.4	1464	Abu54471	Human tum
24	2755	59.4	1464	Abri47417	Breast ca
25	2755	59.4	1464	Abri92064	Human cer

26	2755	59.4	1464	7	ADD14142	Human src
27	2755	59.4	1464	7	ADP65246	Human alp
28	2755	59.4	1464	8	ADQ19470	Human sof
29	2755	59.4	1464	8	ADQ29653	Human col
30	2752.5	59.3	1463	4	AAE02532	Bovine al
31	2752	59.3	1464	7	AD87048	Human pan
32	2750	59.3	1107	2	AAR89472	Collagen/
33	2750	59.3	1169	2	AAR89469	Collagen/
34	2750	59.3	1169	3	AAY84537	Amino aci
35	2750	59.3	1171	2	AAR89470	Collagen/
36	2750	59.3	1388	2	AAR89471	Collagen/
37	2748.5	59.2	1453	7	ADD45053	Rat Prote
38	2748.5	59.2	1453	7	ADD45057	Rat Prote
39	2748.5	59.2	1453	7	ADD48341	Rat Prote
40	2748.5	59.2	1453	7	ADD45049	Rat Prote
41	2748.5	59.2	1453	7	ADD48337	Rat Prote
42	2748.5	59.2	1453	7	ADD48345	Rat Prote
43	2732.5	58.9	1453	5	ABG93948	Mouse pol
44	2725	58.7	1536	7	AD87051	Human pan
45	2723.5	58.7	1411	3	AAY56800	Human pre

ALIGNMENTS

RESULT 1

ADM48392
ID ADM48392 standard; protein; 821 AA.

XX AC ADM48392;

XX DT 03-JUN-2004 (first entry)

XX DE Human recombinant gelatin-like polypeptide Hu-4.

XX KW Plasma substitute; Gelatin-like protein; plasma expander; human.

XX OS Homo sapiens.

XX PN EPI398324-A1.

XX PD 17-MAR-2004.

XX PF 11-SEP-2002; 2002EP-00078745.

XX PR 11-SEP-2002; 2002EP-00078745.

XX PA (FUJF) FUJI PHOTO FILM BV.

XX PI Bouwstra JB, Toda Y;

XX DR WPI; 2004-229415/22.

PT Composition useful as substitute for plasma, comprises solution of saline and recombinant gelatin-like protein having colloid osmotic function.

Example 1; SEQ ID NO 3; 31pp; English.

The present sequence is the protein sequence of recombinant gelatin-like protein Hu-3. This is a tetramer of human recombinant gelatin-like plasma comprises a solution of saline and a protein having a colloid osmotic function. The protein is a recombinant gelatin-like protein with a molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric point of less than 8. It is especially Hu-1 or Hu-deam ADM48393. The recombinant gelatin-like protein, or a dimer, trimer or tetramer of the protein, is useful as a plasma expander that has a lower clearance rate from blood circulation, provides better and predictable regulation of clearance rate and which is less susceptible to proteolytic degradation than presently used gelatin derivatives. Recombinant gelatin-like proteins that are in essence free of hydroxyproline do not give rise to an immunological reaction with blood samples containing IgE antibodies. The gelatin-like proteins can be covalently attached to

pharmacologically active compounds. After administration, the coupled medicament will not diffuse from the circulating blood into the interstitium. Clearance by liver and kidney will be kept to a minimum, ensuring a more constant plasma level of the medicament. Suitable medicaments include those involved in intervening blood clotting, vasodilation, function of erythrocytes, thrombocytes and leukocytes, immune responses, and blood levels of messenger molecules such as hormones.

Query Match 100.0%; Score 4640; DB 8; Length 821;
Best Local Similarity 100.0%; Pred. No. 3.2e-278;
Matches 821; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GPPGEPGTGLPPGPERGGPGSRGFPQADGVAGPKGPAGERGSPGAPKSGPCEAGR	60
Db	1	GPPGEPGTGLPPGPERGGPGSRGFPQADGVAGPKGPAGERGSPGAPKSGPCEAGR	60
QY	61	GEAGLPGAKGLTSGPSGPDGKTGPPGAGODGRPGPPGARGQGVGFPKGA	120
Db	61	GEAGLPGAKGLTSGPSGPDGKTGPPGAGODGRPGPPGARGQGVGFPKGA	120
QY	121	GEPKAGRGVPPGPAVGPAGKDGAGAAQPPGAPAGERGQGPAGSPGQGLPGPA	180
Db	121	GEPKAGRGVPPGPAVGPAGKDGAGAAQPPGAPAGERGQGPAGSPGQGLPGPA	180
QY	181	GPPGEAGKPGQGVPGDLGAPGSPGAGEPGTGLPPGPERGGPSRGPPGADGVAGPK	240
Db	181	GPPGEAGKPGQGVPGDLGAPGSPGAGEPGTGLPPGPERGGPSRGPPGADGVAGPK	240
QY	241	GPAGERGSPGAPKSGPGEAGLPKAGLGTSGPSGPDGKTGPPGAGQDGR	300
Db	241	GPAGERGSPGAPKSGPGEAGLPKAGLGTSGPSGPDGKTGPPGAGQDGR	300
QY	301	GPPGPPGARGQGVGPPGPKGAGECKAGERGVPPGAVGPKDGAGAQGPPGPA	360
Db	301	GPPGPPGARGQGVGPPGPKGAGECKAGERGVPPGAVGPKDGAGAQGPPGPA	360
QY	361	GPAGEGQGPAGSPGFQGLPGPAGPPGEAGTKPEQGVPGDLGAPGSPGAPGPGTGLP	420
Db	361	GPAGEGQGPAGSPGFQGLPGPAGPPGEAGTKPEQGVPGDLGAPGSPGAPGPGTGLP	420
QY	421	GPPCERGSGRPPGADGVAGPKPAGERGSPGAPKSGPGEAGLPKAGL	480
Db	421	GPPCERGSGRPPGADGVAGPKPAGERGSPGAPKSGPGEAGLPKAGL	480
QY	481	GSPGSPDGDGKTGPPGAGODGRPGPPGPGARGQAGVMGFPKGAAGPPKAGRGVP	540
Db	481	GSPGSPDGDGKTGPPGAGODGRPGPPGPGARGQAGVMGFPKGAAGPPKAGRGVP	540
QY	541	GPPGAVGPKDGAGAAQPPGAPGAGERGQGPAGSPGQGLPGPAGPPGAKPGBQ	600
Db	541	GPPGAVGPKDGAGAAQPPGAPGAGERGQGPAGSPGQGLPGPAGPPGAKPGBQ	600
QY	601	GVPDGLGAPGSPGAPGPGTGLPGPPGERGGPSRGFGADGVAGPKGPAGERGSPGPA	660
Db	601	GVPDGLGAPGSPGAPGPGTGLPGPPGERGGPSRGFGADGVAGPKGPAGERGSPGPA	660
QY	661	GPKGSPCEAGRPGEAGLPGKAGLGTSGSPGPDGKTGPPGAGQDGRPGPPGARGQA	720
Db	661	GPKGSPCEAGRPGEAGLPGKAGLGTSGSPGPDGKTGPPGAGQDGRPGPPGARGQA	720
QY	721	GVMGFFGPKGAAGEPKAGERGVPPGPAVGPAGKDGAGAAQPPGAPAGERGQGP	780
Db	721	GVMGFFGPKGAAGEPKAGERGVPPGPAVGPAGKDGAGAAQPPGAPAGERGQGP	780
QY	781	GSPGFQGLPGPAGPCEAGKPGQGVPGDLGAPGSGPAGG	821
Db	781	GSPGFQGLPGPAGPCEAGKPGQGVPGDLGAPGSGPAGG	821

RESULT 2

XX Homo sapiens.
 XX EP992586-A2.
 FN 12-APR-2000.
 PD 07-OCT-1999; 99EP-00119184.
 XX 09-OCT-1998; 98US-00169768.
 PR (USSU) US SURGICAL CORP.
 XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
 PI WPI; 2000-259138/23.
 XX N-PSDB; AAA12503.
 DR Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX Example 10; Fig 39A-E; 260pp; English.
 PS The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents human collagen 1 (alpha1) helical region,
 CC which may be produced using the method of the invention
 SQ Sequence 1057 AA;

Query Match 59.6%; Score 2765; DB 3; Length 1057;
 Best Local Similarity 55.4%; Pred. No. 1.2e-162;
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY 1 GPPGPPGPTGLPGRGPGSGRFGPGADGVAGPKPAGERGSGPGAGPKGSGPGEA--- 57
 DB 18 GPMGSGRGLFPGPGAGPQGFQFPPEPGAGSGPMGPRGPGPGKNGDDGEAGKP 77

QY 58 GRPGAGLP---GAKGLTSGPSGPGP-----DGKTGPPGAGQDGRPGPPGPGAR 105
 DB 78 GRPGRGPPGQAGRLPCTAGLPGMKHGRGFSGLDGAKGDAGPAGPKGPGSGPENGAP 137

QY 106 GQAGVMGFGPKGAGEPKAGER-----GVPPPGAVGPA-----CKDGEAGAO 150
 DB 138 GQMGRGLPGRGRPGAGPAGARGNDGATGAAGPPGTPGAPPGPFPGAVGAKGAGPQ 197

QY 151 GPPGPGAGRGEQGP-----AGSPGFQGLPGPA 180
 DB 198 GPRGSEGGVGRGEPGPGAGAGPAGNPAGDQPGAKGANGAPGIAGAPGFFGARGPS 257

QY 181 GPPGAGKPGGVGPDLDGAP-----GPSGPA-----GEPGPT 213
 DB 258 GPGGPGPPGKNSGEPGAPGSKGDTCAKGPPGVPVQVQPPGAGEGKRGARGEPGPT 317

QY 214 GLPGRGPGRGGPGSGRFGPGADGVAGPKPAGERGSGPGAGPKGSGPGEAGRPGEAGLP 273

DB 318 GLPGRGPGRGGPGSGRFGPGADGVAGPKPAGERGSGPGAGPKGSGPGEAGRPGEAGLP 377
 QY 274 GLTSGPSGSPDPDKTKTGPAGQDGRPGCPGPGARGOAGVMGPPGPKGAAGRPGKAGER 333
 DB 378 GLTSGPSGSPDPDKTKTGPAGQDGRPGCPGPGARGOAGVMGPPGPKGAAGRPGKAGER 437
 QY 334 GVPGRGAVGAGKDGAGAGAGQGGPGPAGPAGERGEGQGPAGSPGFGQLPGPAGPGEAGKP 393
 DB 438 GVPGRGAVGAGKDGAGAGAGQGGPGPAGPAGERGEGQGPAGSPGFGQLPGPAGPGEAGKP 497
 QY 394 GEQGVPGDLGAPSGPAGE-----PGP-----TGLPGPP 423
 DB 498 GEOGVGDLGAPSGPARGERGFPGRGVQCPGPGAGPRGANGAPNDGAKGDAGAPGAP 557
 QY 424 GERGPGSRGPPGADGVAGPKPAGERGSGPGAGPKGSGPGE-----AGRPGEAGLP 474
 DB 558 GSQAGPLQGMGPRGERGAAGLPGPKGDRGDAGPKGADGSGPKGDLGTGTPGPGPAGAP 617
 QY 475 GAK-----GLTSGPSGSPDPDKTKTGPAGQDGRPGCPGPGARGO----- 515
 DB 618 GDKGESFGSPAGPTGARGAPGDRGEPGPPGAGPAGPAGDQPGKAGEPGDGAAGKADA 677

QY 516 -----AGVMGFPKGAAGEPGKAGERGVPPGPA-----VGPAGKDGEGAGAOQPP 561
 DB 678 GPPGPGAGPAPGPGPIGNVGAPGAKGARGSGAGPPGATGTPGAGRVGPPGSGNAGPPGPP 737

QY 562 GPAGP-----AGRGQGPAGSPGFGQLPGPAGPAGPGEAGKPGGQGVPGDLGAPGSGPAGEP 618
 DB 738 GPAGKEGKGPRGTGTGPRGVEGVPGPAGPAGKEGSGADGAGPAGTGTGPGQIAGQR 797

QY 619 GPTGLPGRGPGRGGSGRFGPGADGVAGPKPAGERGSGPAGP---KGSPEGARPGEA 675
 DB 798 GVVGLPQGRGPRGPPGLPGSGEPGKQGPSASGERGPPGPMGPPGLAGPPGSGREGAP 857

QY 676 GLPAGKGLTSGPSGSPDPDKTKTGPAGQDGRPGCPGPGARGOAGVMGFPKGAAGEP 735
 DB 858 GAESGPRGDSGPAKGDRTGTGTPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPV 917

QY 736 GKAGERGVPPGAVGAPGAKDGE-----AGAQGPPGP-----AQPA 771
 DB 918 GPAGARGPAGPQGPGRGKGTGTEGQDRIKGRHFGSLQGGPPGPGSGPGEQGSASGPA 977

QY 772 GEREGQGPAGSP---GFGQLPGPAGPGEAGKPGEQGVPGDLGAPGSGPAG 820
 DB 978 GPRGPPGSGAGAPGKDXGLNLGLPPIGPPGPRGRTGDAGPVGPPGPPGPPGPPG 1029

RESULT 6
 AAY84403
 ID AAY84403 standard; protein; 1058 AA.
 XX
 AC AAY84403;
 XX
 DT 12-JUL-2000 (first entry)
 XX
 DE Amino acid sequence of human type 1 (alpha1) collagen polypeptide.
 XX
 KW Alpha1 collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;
 KW collagen; mussel adhesive protein; bioadhesive.
 XX
 OS Homo sapiens.
 XX
 FN WO200014201-A1.
 XX
 PD 16-MAR-2000.
 XX
 PF 07-SEP-1999; 99WO-US020462.
 XX
 PR 09-SEP-1998; 98US-0099652P.
 XX
 PA (USSU) US SURGICAL CORP.
 PA (PAOL/) PAOLELLA D N.
 PA (GRUS/) GRUSKIN E A.

The specification describes a method for producing an extracellular matrix protein or its fragment. The extracellular matrix protein is capable of self aggregating in a cell which does not ordinarily produce extracellular matrix proteins. The method comprises optimising a nucleic acid sequence for expression in the cell by substitution of codons preferred by that cell for naturally occurring codons not preferred by the cell, incorporating the nucleic acid sequence into the cell, and contacting the cell with a hypertonic growth medium containing at least one amino acid, selected from the group consisting of trans-4-hydroxyproline and 3-hydroxyproline to allow at least one of the amino acids to be assimilated into the cell and incorporated into the extracellular matrix protein. The method may be used to make host cells assimilate and incorporate trans-4-hydroxyproline into proteins. This is especially useful in the recombinant production of proteins such as collagen, fibronectin and fibronectin whose ability to self aggregate and produce functional proteins depends on the post translational hydroxylation of proline. The method is also useful in studying the structure and function of polypeptides which do not normally contain trans-4-hydroxyproline. The present sequence represents a chimeric collagen 1 (alpha1)/decorin protein, which may be produced using the method of the invention

Query Match 59.6%; Score 2765; DB 3; Length 1107;
 Best Local Similarity 55.4%; Pred. No. 1.3e-162;
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

1 GPPGPGTGLPGRGPGSGRPGADGAGVAGPKGAGSGRPGAGPKGSGPGEA--- 57
 18 GPMGSGRGLPGPGAGPQGFQGPGEPEGPGAGPMGPRGPPGPKNGDDGAGKP 77
 58 GRPGEAGLP---GAKGLTSGPSGPGP-----DGKTGPPGAGQGRGPPGPGAR 105
 78 GRPGRGPPGPGQAGRLPCTAGLPGMKHGRGSLDGAGDAGPAGKGEPSGPGENGAP 137
 106 GQAGVMGPPGPKGAAGEPKAGER-----GVPGPPGAVGPA-----GKDGAEAGQ 150
 138 GQMPRGLPGRGRGPGAGARGNDGATGAGPPGTGAGPPGPGVAGKAGGAGPQ 197
 151 GPPGPGAGPAGERGEGP-----AGSPFGQLPGPA 180
 198 GPRGSEGGVGRGPGPGAGAGPAGNPGADGQPGAKGANGAPGIAGPFGARGPS 257
 181 GPPGAGPAGGEGVQVGDLAGP-----GPSGPA-----GEPGPT 213
 258 GPGGPGPPGPKNGSGEPGAPGKGTGAKGPPGVPVQGPAGPAGERGKRGARGEPGT 317
 214 GLPGRGPGRGGPGSGRPGADGAGVAGPKGAGRSGPAGPKGSGPAGRPGAEGLPGAK 273
 318 GLPGRGPGRGGPGSGRPGADGAGVAGPKGAGRSGPAGPKGSGPAGRPGAEGLPGAK 377
 274 GLTSGSGSPGDPGKTGPPGAGQDGRGPPPPGARGQAGVMGPPGKGAAGEPGKAGER 333
 378 GLTSGSGSPGDPGKTGPPGAGQDGRGPPPPGARGQAGVMGPPGKGAAGEPGKAGER 437
 334 GVPGPAGVAGPAGKDGAGAGQGPAGPAGRSGPAGSGPFGQLPGPAGPPGEAGKP 393
 438 GVPGPAGVAGPAGKDGAGAGQGPAGPAGRSGPAGSGPFGQLPGPAGPPGEAGKP 497
 394 GQGVPGDLGAPGSGPAGE-----PGP-----TGLRGPP 423
 498 GQGVPGDLGAPGSGPAGE-----PGP-----TGLRGPP 423
 424 GRRGPGSGRPGADGAGVAGPKGAGRSGPAGPKGSGPGE-----AGRPGEAGLP 474
 558 GSGGAPGLQCMPEGRGAGLPKPGKDRGDAGPKGADGSGPKDGVRLGTGTPGPPAGAP 617
 475 GAK-----GLTSGSGSPGDPGKTGPPGAGQDGRGPPGPPGARGQ----- 515
 618 GPKGSGSGSPGAGTARGAPGDRGEPGPPGAGPAGPPGAGDQGPAGKAGPDAGAKGDA 677
 516 -----AGVMGPPGPKGAAGEPKAGRSGPAGP-----VGPAGKDGAEAGQPP 561

678 GPPGPGAGPAGPPGPIGNVAGAPKAGKARGSGAGPCATGTPGGAAGRVGPPSPGAGPPGPP 737
 562 GPAGP---AGERGQAGAGSPGFGQLPGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 618
 738 GPAGKGGKGRGRTGTGTPAGRPGEVGPPGPPGAGKESGADGAGPAGPAGTGTGPGQIAGQR 797
 619 GPTGLPGRGPGRGGPGSGRPGADGAGVAGPKGAGRSGPAGPAGPAGPAGPAGPAGPAGP 675
 798 GVVGLPQRGGRGPPGPGSGRPGKQPSGASGERGPPGPMGPPGLAGPFGSGREGAP 857
 676 GLPGKAGLTGSPGSGPDPGKTGPPGAGQDGRGPPGPPGARGQAGVMGPPGPKGAAGEP 735
 858 GAEGSPGRDGSAGKGRGTGTGTPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 917
 736 GKAGRGVPPGPPGAVGAPKADGE-----AGAGGPPG-----AGPA 771
 918 GPAGARGPAGPQGRGDKGTGEGQDGRGKGRHGFGLQGPSPGSGPAGSGAGSAGPA 977
 772 GERGQAGPAGSP---GFGGLPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 820
 978 GPRGPPGAGAPGKDGKGLNGLPGIPGPRGRTGDAGPVGPPGPPGPPGPPG 1029

RESULT 8
 AAY84538
 ID AAY84538 standard; protein; 1171 AA.
 XX
 AC AAY84538;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE A chimeric collagen 1 (alpha1)/TGF-beta1 protein.
 XX
 KW Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibronectin; fibronectin; post translational hydroxylation;
 KW ss. transforming growth factor-beta1; TGF-beta1; chimera.
 XX
 OS Homo sapiens.
 OS Unidentified.
 OS Chimeric.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 858 /note= "Gly encoded by GCT"
 FT
 XX
 FN EP992586-A2.
 XX
 PD 12-APR-2000.
 XX
 PF 07-OCT-1999; 99EP-00119184.
 XX
 PR 09-OCT-1998; 98US-00169768.
 XX
 PA (USU) US SURGICAL CORP.
 XX
 XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
 PI N-PSDB; AAA12498.
 DR WPI; 2000-259138/23.
 XX
 DR N-PSDB; AAA12498.
 XX
 PT Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX
 PS Claim 23; Fig 15; 260pp; English.
 XX
 CC The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC produce extracellular matrix proteins. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell,
 CC incorporating the nucleic acid sequence into the cell, and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibronectin and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a chimeric collagen 1 (alpha1)/decorin
 CC protein, which may be produced using the method of the invention

incorporating the nucleic acid sequence into the cell; and contacting the cell with a hypertonic growth medium containing at least one amino acid, selected from the group consisting of trans-4-hydroxyproline and 3-hydroxyproline to allow at least one of the amino acids to be assimilated into the cell and incorporated into the extracellular matrix protein. The method may be used to make host cells assimilate and incorporate trans-4-hydroxyproline into proteins. This is especially useful in the recombinant production of proteins such as collagen, fibrinogen and fibronectin whose ability to self aggregate and produce functional proteins depends on the post translational hydroxylation of proline. The method is also useful in studying the structure and function of polypeptides which do not normally contain trans-4-hydroxyproline. The present sequence represents chimeric collagen 1 (alpha1)/transforming growth factor-beta1 (TGF-beta1) protein, which may be produced using the method of the invention

XX SQ Sequence 1171 AA;

Query Match 59.6%; Score 2765; DB 3; Length 1171;
 Best Local Similarity 55.4%; Pred. No. 1.3e-162;
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY	1	GPFGPGTGLPDPGPPGGRGPGSRGPGADGVAGPKGPGAGRGSPGAGKSPGEA---	57
DB	18	GPWGSPGRLGPPGAGPGQGGPGGEGEGAGSPGMPGPPGPKNGDDGAGKP	77
QY	58	GRFGAELP---GAKGLTSGSPGSP-----DKTGPDPGAGQDGRPPGPPGAR	105
DB	78	GRFGGPPGPPGQAGRLGPGTAGLPGMKHGRFSGLDGAKGADGAPGPKGPPGNGAP	137
QY	106	GQAGVMGFPKGAAGSPGKAGER-----GVPPGAVGPA-----GKDGAGAQ	150
DB	138	GQMGPRGLPGRGPGAPGARGNDGATGAAGPPGPTGPPAGPPGAVGAKGAGPQ	197
QY	151	GPPGAPGAGERGQGP-----AGSPGFQGLGPA	180
DB	198	GPRGSEGPQVGRGPPGPPGAGAGPAGNPDGQPGKAGKANGAPGAGPPGARGPS	257
QY	181	GPFGAEGKPGQVFDLGNP-----GPSGPA-----GEPGPT	213
DB	258	GPGGPGPPGPGKNSGEPGAGSKGTGAKGEPGVGVQPPGAGEGKRGARGEPGT	317
QY	214	GLPFPGERGPGSRGFPAGDGVAGPKGPGAGRGSPGAPGKSGPGEAGPGEAGLPGAK	273
DB	318	GLFPGERGPGSRGFPAGDGVAGPKGPGAGRGSPGAPGKSGPGEAGPGEAGLPGAK	377
QY	274	GLTSGSPGSPDGTGPPGAGQDGRPPGPPGARGQAGVMGFPKGAAGEPGKAGER	333
DB	378	GLTSGSPGSPDGTGPPGAGQDGRPPGPPGARGQAGVMGFPKGAAGEPGKAGER	437
QY	334	GVPPPGAVGAPGAKDGEAGAGQPPGAPGAGRGEGGPGAGSGFGQGLPGPAGPPGAGKP	393
DB	438	GVPPPGAVGAPGAKDGEAGAGQPPGAPGAGRGEGGPGAGSGFGQGLPGPAGPPGAGKP	497
QY	394	GEQGVFDLGAAGSPGAGE-----PGP-----TGLPGPP	423
DB	498	GEQGVFDLGAAGSPGAGERGPPGPPGAGRGANGANGPNDGAKDAGAPGAP	557
QY	424	GRGGPGSRGFPAGDGVAGPKGPGAGRGSPGAPGKSGPGE-----AGRPGEAGLP	474
DB	558	GSQAGFLQMGFERGAAGLPGKDRGDAGPKGADGSGDKGVRLGTGPIGPPGAPAG	617
QY	475	GAK-----GLTSGSPGSPDGTGPPGAGQDGRPPGPPGARGQ-----	515
DB	618	GDKGSGSPGAPGTGARGAGDRGEPGPPGAGFPAGPPGADGQPGKAGEPQDAGAKGDA	677
QY	516	-----AGVMGFPKGAAGEPGKAGERGVPPGPA-----VCPAGKDGAGAGQAPP	561
DB	678	GPPGAPGAPGPPGPIGNVAPKAKGARGAGPPGATGFFGAGRVGPPGSGNAGPPGPP	737
QY	562	GPAGP-----AGEGEGPAGSPFQGLPGPAGPPGAGKPGEGQVDPDGLGAPGSPGAGE	618
DB	738	GPAGKGGKGRGETGTPAGRPGEVPPGPPGAGEKSGPGADGAPGAGFTGPGQAGQR	797

QY	619	GPTGLPGPBERGPGSRGFPAGDGVAGPKGPGAGRGSPGAPG---	KGSPGAGRPGEA	675
DB	798	GVVGLPGGGERGFPGLFGPSGEPKQKQPSGASGERGPPGPPGLAGPPGESGRREGAP		857
QY	676	GLPFGAKGLTSGSPGPPDKTGPAGQDGRPPGPPGARGQAGVMGFPKGAAGEP		735
DB	858	GAEGSPGRDGSFGAKGDRGETGPAGFPAGAPGAPGVPAGKSGDRGETGPAGPAGPV		917
QY	736	GKAGERGVPPGAVGAPKAGDE-----AGAQQPPGPP-----AGPA		771
DB	918	GPAGARGPAGQPPGDRGKGETGEGQDGRGKGRHGFSGLQPPGPPGSPGEGQPSGASGPA		977
QY	772	GERGQGPAGSP---GFGQLPGPAGPCEAGKPGEGQVPGDLAGPSPGSPAG		820
DB	978	GPRGPPGSAGAPGKDLNGLFGPIGPPGPRGRTGDAGVPVPPGPPGPPG		1029

RESULT 9
 AAW68485

ID AAW68485 standard; protein; 1464 AA.

XX AC AAW68485;

XX DT 08-DEC-1998 (first entry)

XX XX Human recombinant collagen protein.

XX KW Primer; PCR; amplification; human; collagen; mammal; plant; prosthesis;
 KW cardiac valve; ligament; tendon; skin; gingival implant; perfumes;
 KW nerve regeneration; antibiotic; growth factor; cancer; inflammatory;
 KW gelatin; glue; food.

XX OS Synthetic.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..22

FT Protein /note= "signal peptide"

FT Protein /note= "mature protein"

FT Cleavage-site 161

FT Cleavage-site /note= "cleavage site for aminopeptidase"

FT Cleavage-site 1218

FT Cleavage-site /note= "cleavage site for carboxypeptidase"

XX PN W09827202-A1.

XX PD 25-JUN-1998.

XX PF 17-DEC-1997; 97WO-FR002331.

XX PR 17-DEC-1996; 96FR-00016224.

XX PA (BIOC-) BIOCEM SA.

XX PI Gruber V, Exposito J, Ruggiero F, Conte J, Garrone R, Merot B;

XX PI Bournat P;

XX DR WPI; 1998-362771/31.

XX DR N-PSDE; AAW60814.

XX XX New recombinant nucleic acid for expressing collagen or derivatives in plants - useful as, e.g. bio-materials and in therapeutic, cosmetic and odontological compositions.

XX PS Disclosure; Fig 7; 138pp; French.

XX CC The invention relates to the production of mammalian collagen in plants. 2 clones: alpha3 and alpha22, spanning the human collagen type I gene were isolated from a MG-63 osteosarcoma library. Clone alpha3 contained 83 bp of the 5' untranslated region and the first 1920 bp of coding sequence, whereas clone alpha22 contained sequence encoding amino acids

171-1454 of the protein and around 500 bp of the 3' untranslated region. The 2 clones were used to generate a number of fragments which were used to construct composite sequences encoding variant collagen molecules. The fragments are: (A) containing nucleotides (nt) -4 to 479; (B) containing TAA upstream of the sequence encoding the PDS (pathogenesis-related protein S) signal peptide and bases 66-77 from the sequence encoding the N-terminus of the pro-collagen amino propeptide domain; (C) the whole of the amino propeptide domain (nt 72-479); (D) all of the amino-telopeptide domain (nt 474-534) and the N-terminus of the helical region (nt 535-1920); (E) the DralII-BamHI fragment (1709-2808) of alpha22, encoding aa 567-936 of the central helical domain; (F) the BamHI-EcoRI (2803-4362) region of alpha22, encoding aa 936-1192 in the central helical domain and aa 1193-1454 in the C-propeptide domain; (G) the C-terminus of the C-propeptide domain (aa 1346-1464) plus stop codons, and (H) as G but encoding aa 1343-1401 and also including the KDEL motif for retention in the ER. This sequence represents a recombinant human collagen. The encoding gene was constructed from fragments (A), (D), (E), (F) and (G). The recombinant gene is used for expression of mammalian collagen in plant cells. The transformed plants, their extracts and parts are useful as biomaterials (haemostatic compresses, sponges or bandages) and in pharmaceutical, medical, odontological, cosmetic and biotechnological compositions (e.g. as prostheses for cardiac valves, ligaments or tendons; skin substitutes; gingival implants; microcapsules for perfumes; guide tubes for nerve regeneration; slow release products for antibiotics, growth factors, anticancer agents or anti-inflammatories; surgical thread and components of ointments). They are suitable for treating any disorder related to collagen dysfunction and gelatin, produced from collagen, is used to produce glues, surgical prostheses and foods

Query Match 59.6%; Score 2765; DB 2; Length 1464;
 Best Local Similarity 55.4%; Pred. No. 1.6e-162;
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

1 GPPGPGTGLPPGRRGGGSGRFGPCADGVAGPKGAGRGSPGAPGKSGPCEA--- 57
 179 GPMGSGRGLPPGAPGPGQGFQGPGEPEGASGPMGPRGPPGPKNGDDGAGKP 238
 58 GRPGAGLP---GAKGLTSGSPGPGP-----DGKTGPPGAGQDGRPGPPGPGAR 105
 239 GRPGRRGPPGQAGRLPGTAGLPMKMGHRGFGSLDGAKDAGPAGPKGFGSPGNGAP 298
 106 GQAGVMGPPGPKGAAGEGPKAGER-----GVPPGPGAVGPA-----GKGGEAGQ 150
 299 GQMCPRLGLGRRGPGAPGAPGARGNDGATGAAGPPGTPGAGPPGPGCAVGAKEAGPQ 358
 151 GPPGAPGAGRGEGP-----AGSPGFOGLPGPA 180
 359 GPRSEGGVGRGPPGPGAGAGPAGNPGADGQPGAKGANGAPGIAGAPGPGARGPS 418
 181 GPPGAGKPGGQGVPGDLGAP-----GPSGPA-----GEPGPT 213
 419 GPQPGGPPGKNGSEFGAPSGKDTGAKGPPGVQGPFGPAGGEGKRGARPEPT 478
 214 GLPGPPGRRGSGRFFPGADGVAGPKGAGRGSPGAPGKSGPGEAGRPCEAGLPKAK 273
 479 GLPGPPGRRGSGRFFPGADGVAGPKGAGRGSPGAPGKSGPGEAGRPCEAGLPKAK 538
 274 GLTSGSPGPDGKTGTPGAPGQDGRPPGPPGARGQAGVMGPPGKGAAGEPKAGER 333
 539 GLTSGSPGPDGKTGTPGAPGQDGRPPGPPGARGQAGVMGPPGKGAAGEPKAGER 598
 334 GVPGPPGAVGAPGKDEAGAGQPPGAPGAGRRGQPGAGSPGFOGLPGAPGPGCEAGKP 393
 599 GVPGPPGAVGAPGKDEAGAGQPPGAPGAGRRGQPGAGSPGFOGLPGAPGPGCEAGKP 658
 394 GEQGVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
 659 GEQGVPGDLGAPGSPGAGERGFPGERGVQGPFGPAGPRGANGAPGNDGAKGADGAPAP 718
 424 GERGPGSRGPPGADGVAGPKGAPGAGERGSPGAPGKSGPGE-----AGRPCEAGLP 474

719 GSQAPGLQMPGCGAAGLPLGPKDRDAGPKGADGSPGKDGVRGLTGTPIGPPGAPAG 778
 475 GAK-----GLTSGSPGPPDGKTPGCPAGQDGRPPGPPGARGQ----- 515
 779 GDKGESGSPGAPGTGARGAPGDRGEPGPPGAPGAGPGADGQGGAKGEPGDAKAGDA 838
 516 ----AGVMGPPGPKGAAGSPGKAGRGVPGPPCA-----VGPAGKDEAGAGQPP 561
 839 GPPGAPGAPGPPGPIGNVGFAPGAKGARGAGPPGATGPPGAGRVGPPGSGNAGPPGPP 898
 562 GPAGP----AGERGQSPAGSPGFOGLPGAPGPPGCEAGKPGGQGVPGDLGAPGSPGAGEP 618
 899 GPACKEGKGRGRTGTPAGRGVPGPPGPPGAPGKSGPAGADGAPGAPGTPGQGIAGOR 958
 619 GPTGLPDPGRRGSGRFFPGADGVAGPKGAPGAGERGSPGAPG---KSGPGEAGRPCEA 675
 959 GVVGLPQGRGRRGFGPLGSPGSEPKQKQPSGASGERGPPGPMGPPGLAGPPGESREGAP 1018
 676 GLPCKAGLGTSGSPGPPGDKTGTGPPGAPGQDGRPPGPPGARGQAGVMGPPGKGAAGEP 735
 1019 GAEGSPGRDGSFGAKGRGRTGTPAGPPGAPGAPGAPGVPAGKSGDRGETGAPGAPGV 1078
 736 GKAGERGVPPGPPGAVGAPGKDG-----AGAQGPPGP-----AGPA 771
 1079 GPAGARGAPGPPGPPGDKGTGEGQDRGKGRGFGSLGQPPGPPGSPGEGSPGASGPA 1138
 772 GERGQSPGAPSP---GFGGLPAGPAGPGEAGKPGGQGVPGDLGAPGSPGAPG 820
 1139 GPRGPPGSGAPGKDGGLNGLPGPIGPPGRTDAGVPVGPVPPGPPGPPG 1190

RESULT 10
 ADD45059
 ID ADD45059 standard; protein; 1464 AA.
 XX AC ADD45059;
 XX DT 29-JAN-2004 (first entry)
 XX DE Human Protein P02452, SEQ ID NO 10491.
 XX KW Human; pain; neuronal tissue; gene therapy;
 KW spinal segmental nerve injury; chronic constriction injury; CCI;
 KW spared nerve injury; SNI; Chung.
 XX OS Homo sapiens.
 XX PN WO2003016475-A2.
 XX PD 27-FEB-2003.
 XX PF 14-AUG-2002; 2002WO-US025765.
 XX PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX PA (GEO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 XX PI Woolf C, D'urso D, Befort K, Costigan M;
 XX WPI; 2003-268312/26.
 DR GENBANK; P02452.
 PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX Claim 1; Page; 1017pp; English.
 XX The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also

Claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at [ftp.wipo.int/pub/published.pct](http://wipo.int/pub/published.pct) sequences.

Sequence 1464 AA;

Query Match	59.6%	Score 2765;	DB 7;	Length 1464;
Best Local Similarity	55.4%;	Pred. No. 1.6e-162;		
Matches	561;	Conservative 31;	Mismatches 228;	Indels 192; Gaps 19;
QY	1	GPCEPPTGLPCGPPRGGRGGPGAGDVGAPKGPAGRGSPGAGPKGSGPEA---	57	
Db	179	GPMPSPGRGLPGPPGAPGPGQFGPGCEGPEGPGASGFMGRGPPGPKNGDDGEAGKP	238	
QY	58	GRGCEAGLP--GAKGLTGTSPGSCP-----	105	
Db	239	GRGERGPPGQAGRLPCTAGLPCMKGHRFSLDKAKDAGPAGPKGPGSPGNGAP	298	
QY	106	GQAGVMGFPFGKAAGSPGKAGER-----	150	
Db	299	GOMGPRGLGERGRPGAPGPAGARGNDGATCAAGPPGPTGAGPPGPGAVGAKGEAGPQ	358	
QY	151	GPQPAGPAGERGEQPP-----	180	
Db	359	GPGRSEGQVGRGPPGPPGAGAGPAGNPGADGQPGAKGANGAGPTAGAPGFGARGPS	418	
QY	181	GPPEZAGKGEQGVPGDLGAP-----	213	
Db	419	GPQPGGPPGPKNSGEPGAPGSKDGTAKGEGPCPVGVQPPGPAGEEGRGARGEPCT	478	
QY	214	GLRPPGERGGPSRPPGADGVAGPKGPAGRGSPGAPGKSPGEGRPGAGLPGAK	273	
Db	479	GLPFPGERGGPSRPPGADGVAGPKGPAGRGSPGAPGKSPGEGRPGAGLPGAK	538	
QY	274	GLTSGSPGSPDCKTTPGPAGDGRPPGPPGAPGAGQAGVMGFPKPKGAAGPGKAGER	333	
Db	539	GLTSGSPGSPDCKTTPGPAGDGRPPGPPGAPGAGQAGVMGFPKPKGAAGPGKAGER	598	
QY	334	GVPPGPGAVGPAKDGEGAGQPPGPAGPAGERGEQGPAGSPFGQLPGPAGPPGAGKP	393	
Db	599	GVPPGPGAVGPAKDGEGAGQPPGPAGPAGERGEQGPAGSPFGQLPGPAGPPGAGKP	658	
QY	394	GEGVPGDILGAPGSPGAGE-----	423	
Db	659	GEGVPGDILGAPGSPGARGERGPPRGVQPPGPAGPRGANGAPGNDGAKGDAGAPGAP	718	
QY	424	GERGGPSRGFPADGVAGPKGPAGRGSPGAPKGSPEG-----	474	
Db	719	GSQCAFELQMPGERGAAGLPGPKGDRGDAGPGADGSPKGVGRGLTGPFGPPGAGAP	778	
QY	475	GAK-----GLTSGSPGSPDGKTGPPGAPGQDGRPGPPGARGO-----	515	

Db	779	GDKGESGSPAGPTGARGAGCDGRGEPCCPPGPAFGAGPAGCDGQFGAKGEPGDGAKGDA	838
QY	516	-----AGVMGPPGPKAAGEGPKAGERGVPPGCA-----VGPAGKDGEAGAGPP	561
Db	839	GPPGPAGPAGPPGPGINVGAPCAKAGKSGAGPPGATGFGAAGKRVGPPGSGNAGPPGPP	898
QY	562	GPAGP----AGRBGQGPAGSPFGQLGPPAGPPGGEAGKPGEQGVGPDGLGAFPGSGPAGEP	618
Db	899	GPAGKEGKGPRGETGPAGRPGVEGPPGPPGAGEKSGPGADGPAGAPGTPGQGIAGQR	958
QY	619	GPTGLPGPPGRCGPGSGRPGCADGVAGPKGPAGERSGPPAGP---KGSFGBAGRPGEA	675
Db	959	GVVGLPQORGERGFPGLPGPSGEPKGKPSGASGERGPPGPMGPPGLAGPPGSGREGAP	1018
QY	676	GLPAKGLTSGSPSGPDGKTGTPGPPAGQDGRCPGPPGPCARGCQAVMGPPGPKGAGEP	735
Db	1019	GAESGPRDGSFGAKGDRGHTGAPGPGAPCAFGPVGPAKSGDRGTGAGPAGPV	1078
QY	736	GKAGERGVPPGPGAVGPAKGDE-----AGAQQPPGP-----AGPA	771
Db	1079	GPACARGPAGPQGRGDKGTGREGDRIKGHRGFSGLQGPFGPPGSGQPSGASGPA	1138
QY	772	GERGECQAGSP---GFQGLPGPAGPPGGEAGKPGEQGVCDLGAPGSGPAG	820
Db	1139	GPRGPPGSGAGPKDGLNGLPGIIGPPGPRGRTCDAGPVGPPGPPGPPGPG	1190

RESULT 11
ADD45055
ID ADD45055 standard; protein; 1464 AA.

DT 29-JÁN-2004 (first entry)

DE Human Protein P02452, SEQ ID NO 10487.

KW Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.

OS Homo sapiens.

PN WO2003016475-A2.

PD 27-FEB-2003.

14-AUG-2002; 2002WO-US025765.

PR 14-AUG-2001; 2001US-0312147P.

PR 26-NOV-2001: 2001US-0333347P.

PA (GEHO) GEN HOSPITAL CORP.

XX
FA (FARB) / BAYER AG.XX
PI WOOLF C, D'urso

DR WPI; 2003-268312/26.
DR GENBANK: D02452

XX
DT
New commodityPT preparing a medi
vv

PS Claim 1; Page; 1017pp; English.

The invention discloses a compo

derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent

CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a human protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1464 AA;

Query Match 59.6%; Score 2765; DB 7; Length 1464;
 Best Local Similarity 55.4%; Pred. No. 1.6e-162;
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;
 QY 1 GPPGEPGTGLPFPGRGPGSGRFPFGADGAGVAGPKGAGPAGSGSPGAGPAGSGPGEA--- 57
 DB 179 GPMGSGRGLPFPGRGPGSGRFPFGADGAGVAGPKGAGPAGSGSPGAGPAGSGPGEA--- 238
 QY 58 GRPGEAGLP---GAKLTGSPGSGPDP-----DGKTGPPGAGQDGRGPPGPPGAR 105
 DB 239 GRPGERGPPGQAGRLPGLTGLPGLMKGHRGFGSLDGAKGADGAGPAGPKGEGPSPGAGP 298
 QY 106 GOAGVMGPPGPKGAGEPKKAGER-----GVPPGPGAVGA-----GKGGEAGAQ 150
 DB 299 GQMGRGLPGRGPPGAGPAGARGNDGATGAAGPPGTPGAGPPGPGVAGKAGEAGPQ 358
 QY 151 GPPGAGPAGGERGEGP-----AGSPGFQGLPGLPA 180
 DB 359 GPRSEGGVGRGPPGPPGAGAGPAGNPGADGQPKAGKANGAPGAGPAGPAGPAGPAGP 418
 QY 181 GPPGAGKPGGQVPGDILGAP-----GPSGA-----GEPGPT 213
 DB 419 GPQGGPFPKNSGSEFGAPGSGKDTGAKGPPGVQVQPPGAGPAGPAGPAGPAGPAGP 478
 QY 214 GLPGRGPPGPPGSGRFPFGADGAGVAGPKGAGPAGSGSPGAGPAGSGPAGPAGPAGP 273
 DB 479 GLPGRGPPGPPGSGRFPFGADGAGVAGPKGAGPAGSGSPGAGPAGSGPAGPAGPAGP 538
 QY 274 GLTSPGSGPDPGKTGPPGAGQDGRGPPGPPGAGQAGVMGPPGPKGAAAGEPKKAGER 333
 DB 539 GLTSPGSGPDPGKTGPPGAGQDGRGPPGPPGAGQAGVMGPPGPKGAAAGEPKKAGER 598
 QY 334 GVPGPAGVAGPAGKDEAGAGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 393
 DB 599 GVPGPAGVAGPAGKDEAGAGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 658
 QY 394 GEQVPGDLGAPGSPGAGE-----PGP-----TCLPGRP 423
 DB 659 GEQVPGDLGAPGSPGAGE-----PGP-----TCLPGRP 718
 QY 424 GRRGPPGSGRFPFGADGAGVAGPKGAGPAGSGSPGAGPAGSGPGE-----AGRPCEAGLP 474
 DB 719 GSQAGPGLQMPGERGAGLPGPKDGRDAGPKGADGSGFKDGVKGLTGTPTGPPGAGPAG 778
 QY 475 GAK-----GLTSGSPGSPGDKTGPAGQDGRGPPGPPGPPGARGQ----- 515
 DB 779 GKGSGSPGPPAGTARGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 838
 QY 516 -----AGVMGPPGPKGAGEPKKAGERGVPPGPPGA-----VGPAGKGEAGAGQPP 561

DB 839 GPPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 898
 QY 562 GPAGP-----AGERGGPAGSPGPGGLPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 618
 DB 899 GPAGKGGKGRGRTGTGPRGPGVGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 958
 QY 619 GPTGLPGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 675
 DB 959 GVVGLPQGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1018
 QY 676 GLPAGKGLTSPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 735
 DB 1019 GAEGSPGRDGSPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1078
 QY 736 GKAGERGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 771
 DB 1079 GPAGARGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1138
 QY 772 GERCEQGPAGSP---GFQGLPGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 820
 DB 1139 GPRGPPGAGAPGKDGKGLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1190

RESULT 12
 ADD45051
 ID ADD45051 standard; protein; 1464 AA.
 XX
 AC ADD45051;
 DT 29-JAN-2004 (first entry)
 XX
 DE Human Protein P02452, SEQ ID NO 10483.
 KW Human; pain; neuronal tissue; gene therapy;
 KW spinal segmental nerve injury; chronic constriction injury; CCI;
 KW spared nerve injury; SNI; Chung.
 XX
 OS Homo sapiens.
 XX
 FN WO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX
 PF 14-AUG-2002; 2002WO-US025765.
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 XX (GSHO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 XX
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX
 XX WPI; 2003-268312/26.
 DR GENBANK; P02452.
 XX
 PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1; Page; 1017pp; English.
 CC
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially

expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at <http://wipo.int/pub/published.pct> sequences.

Sequence 1464 AA;

Query Match 59.6%; Score 2765; DB 7; Length 1464;
Best Local Similarity 55.4%; Pred. No. 1.6e-162;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

QY	1	GPPEBGPPTGLPFPGERGGPSRFFPGADGVAGKPGPAGERSPGPAGKPSGPEA----	57
Db	179	GPWSPSGRGLPFPFCAPFPQCFQFPPEGPCEGASGFMGPRGPPGPGKNGDDGEAGKP	238
QY	58	GRPEAGLPE--GAGLGTSGSPSGP-----DGKTGPPPAGQDGRPGPPPGAR	105
Db	239	GLPGERGPPGQAGRLPCTAGLPWKGHRFGSLDGAKGDAGPAGKPEFSGPGENGAP	298
QY	106	QOAGVMGFPKPGAAGEFKAGER-----GVPPFPGAVGPA-----GKDXGAGAG	150
Db	299	GQMGPRGLPGERGRPCADCPAGAGNDGATCAAGFPPTGTGAPGPPFPGAAGKAGEAGQ	358
QY	151	GPFPAGPAGERGEOGP-----ASPPGQGLPGPA	180
Db	359	GRGSEGGVGRGEFPPGPAGAGPAGNPGADQCPGAKGANGAPGTAGPFPGARGPS	418
QY	181	GPPEAGKPGEOQVPCDLGAP-----GPSGA-----GBPGPT	213
Db	419	GQGPCGPPGKNSGSEPGAFCSGKDTCAKECPGVQPPGAPGAEKRGARGEPGT	478
QY	214	GLPBPBGRGGPSRFFPGADGVAGKPGPAGERSGPGAPKPSGPEAGRPCEAGLPGAK	273
Db	479	GLPFPBGRGGPSRFFPGADGVAGKPGPAGERSGPGAPKPSGPEAGRPCEAGLPGAK	538
QY	274	GLTSGSGPDPDKTTPGPBAGQDGRPCPPGPARGQAGVMGFPFGKNAAGEPKKAGER	333
Db	539	GLTSGSGPDPDKTTPGPBAGQDGRPCPPGPARGQAGVMGFPFGKNAAGEPKKAGER	598
QY	334	GVPPPPGAVGPAKDXGAGAGPFPAGPAGERGEOGPAGSPFQGLGPPAGPPGEAGKP	393
Db	599	GVPPFPGAVGPAKDXGAGAGPFPAGPAGERGEOGPAGSPFQGLGPPAGPPGEAGKP	658
QY	394	GSQGVPGDLGAPSGPAGE-----PGP-----TGLPGPP	423
Db	659	GEQGVPGDLGAPSGGARGERFPERGVOGPPGPAGFRGANGAPNDGAKGDGAPGAP	718
QY	424	GERGFGSRGPPGADGVAGKPGPAGERSGPGAPKPSGPE-----ACRPCEAGLP	474
Db	719	GSQAGPLQSMFEGERGAGLPGPKDGRDAGPKADGSGFKDGVRLGTGIPGPPGAPNP	778
QY	475	GAK-----GLTSGSGSPDGKTPPGPAGQDGRPGPPPGARQO-----	515
Db	779	GDKBSGSPGAPGTGARGAPGDRGEFPGPAGFAGPPGADQCPGAKGEPDGAKGDA	838
QY	516	-----AGVMGFPKGAAGEPKAGERGVPCPA-----VCPAGKDXGAGAGPP	561
Db	839	GPPGAPGAPPGPIGNVGAFCAKARGSGPPGATGFPGAARGRVPPGPSGNAGPPPPP	898
QY	562	GPAGP-----AGRBEGQGPAGSPGFQGLPFGAPPPCEAGKPEQGVPPDLAGPSPGAPGP	618

Db	399	GPAGEKGKGRBGTGTGAPRGEVVGPPGPGPAGEKSGPADGPAGAPGTGPGQGIAGQR	958
QY	619	GFTGLPGPPGERGCGPSRGPGADGVAGPKPGACERSGSPGAG- ---KSGSPAGRRPGEA	675
Db	959	GVVGLPQORGERGFPGLPGPSGEPKGQGPSASGERGPPGPMGPPGLAGLPPGSGREGAP	1018
QY	676	GLPAAKGLTGSPSGPGDDKGTGPPGPAQOGRPGPPQPGPARGQAQVWVGPPGPKGAAGEP	735
Db	1019	GAEGSPGRDPSGPAKGRDGTGTGAPGPGAPCAPGVPVPGAKSGDRGTGTGAPGAPGV	1078
QY	736	GKAGERVWPPGPAVGPAAGKDB- ---AGAQGPPGP- ---AGPA	771
Db	1079	GPAGARGPAGPQGRGDKGTGTGQQRDEGIKCHRGFSLGQPPGPGSPGQSPSGAGPA	1138
QY	772	GERGEQAPSP- ---GFQGLPGPAGPPGEAGKPGEGQVPGDLGAPGSGPAG	820
Db	1139	GPGRPPGSAGAPKQDGLNGLPGIIGPPRGRGTGDAGVPVPGPPGPPGPPG	1190

RESULT 13

ADRI16800
ID ADRI16800 standard; protein; 1464 AA.

AC ADR16800;

DT 21-OCT-2004 (first entry)

DE Human collagen I alpha 1 (I) chain protein.

XX
KW
T-lymphocyte; AAA; abdominal aortic aneurysm; human; collagen I; therapy.

OS Homo sapiens.

AA
PN
US2004151731-A1.

PD 05-AUG-2004.

04-FEB-2003: 2003US-00357851.

AA
PR 04-FEB-2003: 2003US-00357851.XX
PA (JICH/) JICHA D L.XX Jicha DL:
PIXX
DR
WPI: 2004-561474/54.

AA PT PT PT

PS Disclosure; SEQ ID NO 1; 24pp; English.

The present invention relates to a T-lymphocyte derived from abdominal aortic tissue which are specifically reactive with collagen I, collagen III or their fragments. The invention is useful to prevent or treat an abdominal aortic aneurysm (AAA) or rupture in a mammal. The present sequence is human collagen I alpha1 (I) chain protein. This sequence is used in the invention.

Sequence 1464 AA:

Query Match	59.6%;	Score 2765;	DB 8;	Length 1464;
Best Local Similarity	55.4%;	Pred. No. 1.6e-163;		
Matches 561;	Conservative	31;	Mismatches 228;	
			Indels 192;	Gaps 19

QY 1 GPPGEPGPTGLPGPPGERGGPGSRGFFGADGVAGPKGPAGERGSPGPAGPKGSPGEA--- 57

D_b 179 GPMGPSGRGLPGPPGAPGPQGFQPPGEPGEFGASGPMGPRGPPGPPKNGDDGEAGKP 238

QY 58 GRPGEAGLP---GAKGLTSPGSPGP-----DGKTGPPGPAGQDGRPPGPPGAR 105

Db 239 GRPGERPPGPQGARGLPCTAGLPGMKGHRGFSGLDGAKGDAGPAGPKGEPGSPGENGAP 298


```
QY 106 GOAGVMGPPGPKGAAGEGKAGER-----GVPGPAGVGA-----GKDGEAGQ 150
Db 299 GQMGPRGLGGRGRPGAPGARGNDGATGAAGPGGTGAGPPGPGAVGAKGAGPQ 358
QY 151 GPPGPAGPAGERGEQGP-----AGSPFGQLPGPA 180
Db 359 GPRGSEGVGVRGPPGPPGAGAGPAGNPGADQPGAKGANGAPGTFGPGARGPS 418
QY 181 GPPGAGKPGGQVPGDLGAP-----GPSGPA-----GERPPT 213
Db 419 GPQGGPGGPKGNSGEPGAPGSKGDTGAKGEPGPGVQGGPPGAGEGKRGARPEPT 478
QY 214 GLPDPGRRGGRGSRGPPGADGAVGPKGSPGAGRPGAGRPGAGLPGAK 273
Db 479 GLPDPGRRGGRGSRGPPGADGAVGPKGSPGAGRPGAGRPGAGLPGAK 538
QY 274 GLTSPGSPGDPGKTGTPGPPGAGQDGRPPGPPGARGQAGVMGFFGPKGAAGEGPKAGER 333
Db 539 GLTSPGSPGDPGKTGTPGPPGAGQDGRPPGPPGARGQAGVMGFFGPKGAAGEGPKAGER 598
QY 334 GVPDPGAVGAGKDGEGAGAGQPPGPPGAGPAGERGEQGPAGSPFGQLPGPAGPPGEGAKP 393
Db 599 GVPDPGAVGAGKDGEGAGAGQPPGPPGAGPAGERGEQGPAGSPFGQLPGPAGPPGEGAKP 658
QY 394 GEOGVPGDLGAPGSPGAGE-----RGP-----TGLPGPP 423
Db 659 GEOGVPGDLGAPGSPGAGERGFTGERGVQPPGAPRPGARGANGAPGNDGAKGDAGAPAP 718
QY 424 GERGPPSRGPPGADGAVGPKGSPGAGRPGAGRPGAGLPGAK 474
Db 719 GSQAPGLQMPGREGAAGLPGPKDGRDAGPKGADSGKDGVRGLTGPFGPPAGAP 778
QY 475 GAK-----GLTSPGSPGDPGKTGTPGPPGAGQDGRPPGPPGARGQ-----515
Db 779 GDKGESGSPGAPGTGARGAPGDRGEPGPPGPPGAGPPGADGQPGAKGEPGDAGAKGDA 838
QY 516 -----AGVMGPPGPKGAAGEGKAGERGVCPGGA-----VGPAGKDGEGAGAGPP 561
Db 839 GPPGAPGAPGPPGPIGNVGAAGKARGAGSAGPPGATGPPGAAGRVGPPGSGNAGPPGPP 898
QY 562 GPAGP-----AGRGEQGPAGSPGQGLPGPAGPPEAGKPGGQVPGDLGAPGSPGAPG 618
Db 899 GPAGKEGKGRGTGTGPRGVEVGPFPFPAGEKSGFADGAGPAGTGTGPGIAGOR 958
QY 619 OPTGLPGPPRRGGRGPPGADGAVGPKGSPGAGRPGAGP-----KGSPCEAGRPGEA 675
Db 959 GWVGLPGQGRGERGPPGLPGSPGPKQGPAGSAGRPPGPPGLAGPPGESGREGAP 1018
QY 676 GLPGAAGLTSFGSPGPPGKTGTPGPPGAGQDGRPPGPPGARGQAGVMGFFGPKGAAGEP 735
Db 1019 GAEGSPGRDGSFGAKGDRGTGTPGPPGAPGAPGAPGVPVGPAGSKGDRGTGTPGAPGV 1078
QY 736 GKAGERGVPPGPAAGVAPGAKDGE-----AGAQQPPGP-----AGPA 771
Db 1079 GPAGARGPAGQPPGDKGTGEGQDRIKGRHGFGLQGGPPGPPGSGEGOGPSGASGPA 1138
QY 772 GERGEQGPAGSP-----GFGQLPGPAGPPEAGKPGGQVPGDLGAPGSPGAPG 820
Db 1139 GPRGPPGSAAGPKDGLNGLGPIGPPGPRGTGDAGVPGPPGPPGPPG 1190
```

RESULT 14

```
ADRI16425
ID ADRI16425 standard; protein; 1464 AA.
XX AC ADRI16425;
XX XX
XX DT 21-OCT-2004 (first entry)
XX DE
XX KW Human collagen I alpha1 (I) chain protein.
XX T-lymphocyte; AAA; abdominal aortic aneurysm; vaccine; therapy; human;
KW collagen I.
```

```
XX Homo sapiens.
OS US2004151732-A1.
FN 05-AUG-2004.
PF 04-FEB-2003; 2003US-00358024.
PR 04-FEB-2003; 2003US-00358024.
PA (JICH/) JICHA D L.
PA (PELU/) PELUSE S.
PI Jicha DL, Peluse S;
XX WPI; 2004-570707/55.
XX Isolated or purified lymphocytes derived from blood in abdominal aortic
XX aneurysm patients useful for early diagnosis of aneurysms are reactive
XX with collagen I, collagen III and/or their fragments.
XX Disclosure; SEQ ID NO 1; 22pp; English.
XX The present provides a T-lymphocyte derived from blood in abdominal
XX aortic aneurysm (AAA) patients which are reactive with collagen I,
XX collagen III and their fragments. The invention is useful in diagnosis,
XX prevention and treatment of abdominal aortic aneurysm or rupture in a
XX mammal. The invention is also useful in the vaccine preparation. The
XX present sequence is human collagen I alpha1 (I) chain protein.
XX Sequence 1464 AA;
QY Query Match 59.6%; Score 2765; DB 8; Length 1464;
Db Best Local Similarity 55.4%; Pred. No. 1.6e-162;
QY Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;
Db 1 GPPPEPPTGLPGRPPGRRGPPGSGRPFQADGAVGPKGAGRSPGPPGPKGSGPGEA---57
179 GPMGPPSGRGLGPPGPPGAPGPGFQGGFPGEPGEPGASGPMGPPGPPGKNGDDGEGAKP 238
QY 58 GRPGEAGLP---GAKGLTSPGSPG---DGTGTPGPPGAGQDGRPPGPPGAP 105
Db 239 GRPGEPPFPQAGARGLPGTAGLPGMKHGRGFSGLDGAAGAGPAGPKGPPGSPGAGP 298
QY 106 GQAGVMGPPGPKGAAGEGPKAGER-----GVPGPAGVGA-----GKDGEAGQ 150
Db 299 GQMGPRGLPGERGPPGAPGAGAGNDGATGAAGPPGTGTPGPPGPPGAVGAKGAGPQ 358
QY 151 GPPGPAGPAGERGEQGP-----AGSPFGQLPGPA 180
Db 359 GPRGSEGVGVRGPPGPPGAGAGPAGNPGADQPGAKGANGAPGTFGPGARGPS 418
QY 181 GPPGAGKPGGQVPGDLGAP-----GPSGPA-----GERPPT 213
Db 419 GPQGGPGGPKGNSGEPGAPGSKGDTGAKGEPGPGVQGGPPGAGEGKRGARPEPT 478
QY 214 GLPDPGRRGGRGSRGPPGADGAVGPKGSPGAGRPGAGRPGAGLPGAK 273
Db 479 GLPDPGRRGGRGSRGPPGADGAVGPKGSPGAGRPGAGRPGAGLPGAK 538
QY 274 GLTSPGSPGDPGKTGTPGPPGAGQDGRPPGPPGARGQAGVMGFFGPKGAAGEGPKAGER 333
Db 539 GLTSPGSPGDPGKTGTPGPPGAGQDGRPPGPPGARGQAGVMGFFGPKGAAGEGPKAGER 598
QY 334 GVPDPGAVGAGKDGEGAGAGQPPGPPGAGPAGERGEQGPAGSPFGQLPGPAGPPGEGAKP 393
Db 599 GVPDPGAVGAGKDGEGAGAGQPPGPPGAGPAGERGEQGPAGSPFGQLPGPAGPPGEGAKP 658
QY 394 GEOGVPGDLGAPGSPGAGE-----RGP-----TGLPGPP 423
Db 659 GEOGVPGDLGAPGSPGAGERGFTGERGVQPPGAPRPGARGANGAPGNDGAKGDAGAPAP 718
```


[illegible]

Search completed: June 17, 2005, 15:08:26
Job time : 140.148 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:46:04 ; Search time 34.8485 Seconds
(without alignments)
1758.668 Million cell updates/sec

Title: US-10-658-989A-3

Perfect score: 4640

Sequence: 1 GPPGEPGPTGLPGERGG.....GEQGVGDLGAPGSPGAGG 821

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/1/iaa/PCUS_COMB.pep:*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2765	59.6	1464	4	US-09-331-347C-21
2	2752	59.3	1461	4	US-09-585-887-9
3	2752	59.3	1461	4	US-09-289-578-9
4	2746	59.2	1057	3	US-08-931-820-1
5	2704	58.3	1341	3	US-08-963-825-18
6	2704	58.3	1341	3	US-09-500-811-18
7	2704	58.3	1341	3	US-09-570-573-18
8	2704	58.3	1341	3	US-09-548-608-18
9	2511	54.1	822	3	US-09-219-849-49
10	2488	53.6	1017	4	US-08-468-996-10
11	2478	53.4	1064	1	US-08-642-285-62
12	2476	53.4	1060	3	US-08-931-820-3
13	2476	53.4	1418	3	US-08-963-825-20
14	2476	53.4	1418	3	US-09-010-999-1
15	2476	53.4	1418	3	US-09-500-811-20
16	2476	53.4	1418	3	US-09-570-573-20
17	2476	53.4	1418	3	US-09-548-608-20
18	2464	53.1	1442	2	US-08-316-650-12
19	2464	53.1	1442	5	PCT-US95-02251-12
20	2370	51.1	1057	3	US-08-931-820-4
21	2350.5	50.7	1078	3	US-08-963-825-21
22	2350.5	50.7	1078	3	US-09-500-811-21
23	2350.5	50.7	1078	3	US-09-570-573-21
24	2350.5	50.7	1078	3	US-09-548-608-21
25	2267.5	48.9	1366	3	US-08-963-825-19
26	2267.5	48.9	1366	3	US-09-500-811-19
27	2267.5	48.9	1366	3	US-09-570-573-19

28	2267.5	48.9	1366	3	US-09-548-608-19
29	2265.5	48.8	1366	4	US-09-585-887-10
30	2265.5	48.8	1366	4	US-09-289-578-10
31	2265.5	48.8	1366	4	US-09-949-016-5882
32	2255.5	48.6	1024	3	US-08-931-820-2
33	2254.5	48.6	1065	1	US-08-642-255-72
34	2254.5	48.6	1065	4	US-09-919-497-56
35	2169	46.7	1806	4	US-09-219-849-5
36	2169	46.7	1806	3	US-08-642-255-53
37	2144.5	46.3	777	1	US-09-219-849-4
38	2144.5	46.2	720	3	US-08-397-633A-50
39	2096	45.2	595	3	US-09-219-849-48
40	2096	45.2	595	3	US-09-219-849-50
41	2057	44.3	1745	4	US-08-795-061-4
42	2055	44.3	837	1	US-08-175-155-68
43	2055	44.3	837	1	US-08-477-509B-103
44	2055	44.3	837	1	US-08-642-255-101
45	2055	44.3	837	2	US-08-707-237A-75

ALIGNMENTS

RESULT 1

US-09-331-347C-21

; Sequence 21, Application US/09331347C

; Patent No. 6617431

; GENERAL INFORMATION:

; APPLICANT: Meristem Therapeutics, S.A.

; APPLICANT: Meristem Therapeutics, S.A.

; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Met

; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Met

; FILE REFERENCE: 1149-3

; CURRENT APPLICATION NUMBER: US/09/331.347C

; CURRENT FILING DATE: 1999-08-17

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 21

; LENGTH: 1464

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-331-347C-21

Query Match 59.6%; Score 2765; DB 4; Length 1464;

Best Local Similarity 55.4%; Pred. No. 3.3e-171;

Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

Qy	1	GPPGEPGPTGLPGERGGPFGADGVAGPKGAGRGSPGAPGKSGPCEA---	57
Db	179	GPMGSPGRGLPGLPPGAPGPGFQGPGEPEGASGMGPRGPPGPKNGDDGEAKP	238
Qy	58	GRPGEAGLP---GAKGLTSGSPGSP-----DGKTGPPGAGODRPPGPPGAP	105
Db	239	GRPGEAGLP---GAKGLTSGSPGSP-----DGKTGPPGAGODRPPGPPGAP	298
Qy	106	COAGVMGPPGPKGAAGSPGKAGER-----GVPPGPAVGPA-----GKGEAGAQ	150
Db	299	QOMGPRGLPGERGRPGAPGAGATGATGATGATGATGATGATGATGATGATGATGAT	358
Qy	151	GPPGPAAGPAGRGQP-----AGSPGFCQLPGPA	180
Db	359	GPRSEGGVGRGPPGPPGAGAGPAGNPGADGQPGAKGANGAPGIAGAPGPPGARGPS	418
Qy	181	GPPGEAGPKGQGVPGDLGAP-----GPSGPA-----GEPGPT	213
Db	419	GPGGPPGPPGKNSGEPGAGSGKDTCAKCEPGVPVQGPVQGPVQGPVQGPVQGPVQGP	478
Qy	214	GLPFPGERGGPGRFPAGDGVAGPKGAGRGSPGAPGKSGPCEAGRPEAGLPGAK	273
Db	479	GLPFPGERGGPGRFPAGDGVAGPKGAGRGSPGAPGKSGPCEAGRPEAGLPGAK	538
Qy	274	GLTSPGSPGPDGKTGPPGAGODRPPGAPGAGQVMGPPGPPGAPGPPGAPGPPGAPGPPG	333

Db	539	GLTGSFGSGPPDGKTTGPPGAGQDGRPPGPPGARGQAGVMGFPFGKGAEGEPKACGER	598
QY	334	GVPGPPGAVGPAKDGCEAGACQPPGAPGAGBRGQGPAGSPGFQGLPKPAGPPGCEAKCP	393
Db	559	GVPGPPGAVGPAKDGCEAGACQPPGAPGAGBRGQGPAGSPGFQGLPKPAGPPGCEAKCP	658
QY	394	GEQVPGDLGAPGPPSGPAGE-----PGP-----TGLPGPP	423
Db	659	GEQVPGDLGAPGPPSGARGERFPGERGVQPPGAPGPRGANGAPGNDCAKGDAGACAP	718
QY	424	GERGPGSRGPPGADGVAGPKGAPACERGSPPGAPKSGPGE-----AGRPGEACLP	474
Db	719	GSQAPGLQMGPFGERGAAGLPGPKDGRDAGPKGADGSPGKDXVGRGLTGPICPPGPAGAP	778
QY	475	GAK-----GLTSGPSGPPDGKTFPPGAPGQDGRPPGPPGARGQ-----	515
Db	779	GDKGESGPPAGPTGARGAPGDRGEPGPPGAPGAPGADGQPKAGEFGPDAGAKGDA	838
QY	516	-----AGVMGPPPKGAAGEPKKAGERVVPDPGA-----VGPAGKDGCEAGACQPP	561
Db	839	GPPGPAGPAGPPGPTIGNVAPCAKARGASGPPGATGFPGAAGRVGPPGPSGNAGPPGPP	898
QY	562	GPAGP---AGBRGQGPAGSPGFQGLPGPAGPPCEAGKPGQGVPGDLGAPGSGGAPAGE	618
Db	899	GPACKEGKGPRGETGPAGRGEVGPFGPPGPAKEKSGFADGPAGAPGTPGPGIGAQOR	958
QY	619	GPTGLPDPGPRGPGSGRFPFGADGVAGPKGAPACERGSPPGAPG---KGSFGEAGRPEEA	675
Db	959	GVVGLPGQGRGRRFPGLPGPSGEPGKQGPSCASGERPPGPMGPPGLAGPPGESREGAP	1018
QY	676	GLPCKAGLGTSGPSGPPDGKTFPPGAPGQDGRPPGPPGARGQAGVMGPPGPKGAAGEP	735
Db	1019	GAEGSPGRDGSFGAKGRDGETGTPGPPGAPGAPGAPGVGPAKSGDRGETGPGAPGAPV	1078
QY	736	GKAGERGVPPGPGVAGPAGKDGE-----AGAQPPGPG-----AGPA	771
Db	1079	GPAGARGPAGPQGRDKGETGETGEQDGRIGKHRRGFSGLQGPFPGPSFGQSGSASGPA	1138
QY	772	GERCEQGPASP---GFGQGLPGPAGPPCEAGKPGCEQGVPGDLGAPGSGPAG	820
Db	1139	GPRGPPGSAGAPKDGKLNGLGFPITGPPGPRGRTDAGVPGPPGPPGPPGPPG	1190

RESULT 2

```

US-09-585-887-9
; Sequence 9, Application US/09585887
; Patent No. 6413742
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/585,887
; CURRENT FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/289,578
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-585-887-9

```

Query Match 59.3%; Score 2752; DB 4; Length 1461;
Best Local Similarity 55.2%; Pred. No. 2.2e-170;

RESULT 3
US-09-289-578-9

```
; Sequence 9, Application US/09289578
; Patent No. 6428978
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/289,578
; CURRENT FILING DATE: 1999-04-10
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-289-578-9

Query Match 59.3%; Score 2752; DB 4; Length 1461;
Best Local Similarity 55.2%; Pred. No. 2.2e-170;
Matches 559; Conservative 31; Mismatches 230; Indels 192; Gaps 19;

QY 1 GPPGPGTGLPGPPGGRGGGSGRFPAGDGVAGPKGAGRGSPGAPGKSGPGEA--- 57
DB 176 GPMGSGRGLPGPPGAGPQGFQFPPEPGEAGSGPMGPRGPPGPKNGDDGEAKP 235
QY 58 GRPGGAGLP---GAKGLTSGSPGPGP-----DGKTGPPGAGQDRPFPPPGAG 105
DB 236 GRPGGPPGPGQAGRLPCTAGLPCMKHGRGSLDGAKGADGAPGAPGKGEFGSPG 295
QY 106 GQAGVMGPPGKGAAGEPKAGER-----GVPGPPGAVGA-----GKDEAG 150
DB 296 GQMGRLGPPGGRGPGAGARGNDGATGAAGPPGTPGAPGPPGPPGAVGAKGE 355
QY 151 GPPGAGPAGRGEGP-----AGSPGQGLPGPA 180
DB 356 GPRGSGGPPGPPGPPGPPGAGAGPAGNADGQFAGKANGAGIAGAPGPPGARG 415
QY 181 GPPGAGPAGRGEGP-----GPPGGA-----GEPGP 213
DB 416 GPGGPPGPPGKNGSGEPGAGKSGDTCAGKEPVPVQGPFGPAGRGKRGARPG 475
QY 214 GLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 273
DB 476 GLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 535
QY 274 GLTSGPSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 333
DB 536 GLTSGPSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 595
QY 334 GVPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 393
DB 596 GVPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 655
QY 394 GQVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 423
DB 656 GQVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 715
QY 424 GERGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 474
DB 716 GSQGAPGLQMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 775
QY 475 GAK-----GLTSGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 515
DB 776 GDKSGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 835
QY 516 -----AGVMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 561
```

```
; Sequence 9, Application US/09289578
; Patent No. 6428978
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/289,578
; CURRENT FILING DATE: 1999-04-10
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-289-578-9

Query Match 59.3%; Score 2752; DB 4; Length 1461;
Best Local Similarity 55.2%; Pred. No. 2.2e-170;
Matches 559; Conservative 31; Mismatches 230; Indels 192; Gaps 19;

QY 1 GPPGPGTGLPGPPGGRGGGSGRFPAGDGVAGPKGAGRGSPGAPGKSGPGEA--- 57
DB 176 GPMGSGRGLPGPPGAGPQGFQFPPEPGEAGSGPMGPRGPPGPKNGDDGEAKP 235
QY 58 GRPGGAGLP---GAKGLTSGSPGPGP-----DGKTGPPGAGQDRPFPPPGAG 105
DB 236 GRPGGPPGPGQAGRLPCTAGLPCMKHGRGSLDGAKGADGAPGAPGKGEFGSPG 295
QY 106 GQAGVMGPPGKGAAGEPKAGER-----GVPGPPGAVGA-----GKDEAG 150
DB 296 GQMGRLGPPGGRGPGAGARGNDGATGAAGPPGTPGAPGPPGPPGAVGAKGE 355
QY 151 GPPGAGPAGRGEGP-----AGSPGQGLPGPA 180
DB 356 GPRGSGGPPGPPGPPGPPGAGAGPAGNADGQFAGKANGAGIAGAPGPPGARG 415
QY 181 GPPGAGPAGRGEGP-----GPPGGA-----GEPGP 213
DB 416 GPGGPPGPPGKNGSGEPGAGKSGDTCAGKEPVPVQGPFGPAGRGKRGARPG 475
QY 214 GLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 273
DB 476 GLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 535
QY 274 GLTSGPSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 333
DB 536 GLTSGPSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 595
QY 334 GVPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 393
DB 596 GVPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 655
QY 394 GQVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 423
DB 656 GQVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 715
QY 424 GERGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 474
DB 716 GSQGAPGLQMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 775
QY 475 GAK-----GLTSGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 515
DB 776 GDKSGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 835
QY 516 -----AGVMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 561
```

```
836 GPPGAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 895
562 GPAGP---AGERGQGPAGSPGFGGLPGPAGPPGPPGPPGPPGPPGPPGPPGPP 618
896 GPAGKGGKGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 955
619 GPTGLPGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 675
956 GVVGLPQGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1015
676 GLPAGKGLTSGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 735
1016 GAEGSPGRDSSPGAKGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1075
736 GKAGERGVPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 771
1076 GPPGARGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1135
772 GERGQGPAGSP---GFGGLPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 820
1136 GPRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1187

RESULT 4
US-08-931-820-1
; Sequence 1, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type I
US-08-931-820-1

Query Match 59.2%; Score 2746; DB 3; Length 1057;
Best Local Similarity 56.3%; Pred. No. 4.1e-170;
Matches 556; Conservative 32; Mismatches 232; Indels 168; Gaps 18;

QY 1 GPPGEP-----GPTGLPGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 45
DB 42 GPPGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 101
QY 46 -----GPAGPK---GSPGAGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 80
DB 102 GMKGRGSGLDGAKGADGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 161
QY 81 ---DKT---GPPGAGQDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 123
DB 162 GNDGATGAAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 221
```

QY	124	GKAGERGVPPGAVGPKADGEAGAAQGP	PPGAPGAGERGQGPAGSGPFOGLPGAGPP	183
DB	222	GPAPNPGADQGPAGKANGAPGIAAP	FPGARGSPQGPGGPTGARGLVGP	281
QY	184	GEACKPEQGVPGDLGAPGPPS----	GPAGEPPTGLPPGGERGGSGRGFPADGVA	237
DB	282	GESCNKEPGSAGPQGP	PPGSGEGRKPNGEASAGPPFGLRG	341
QY	238	GPKPAGERGSPGAPGKSGPGEAGLPG	AKGLTGSFSGPPGDKTGP	297
DB	342	GPKPAGERGSPGAPGKSGPGEAGLPG	AKGLTGSFSGPPGDKTGP	401
QY	298	GRPGPPPGARGQGVMPGPPKGAAGE	PKGAGERGVPPGAVGPAGKDEAGACQPP	357
DB	402	GRPGPPPGARGQGVMPGPPKGAAGE	PKGAGERGVPPGAVGPAGKDEAGACQPP	461
QY	358	GPAPGAGERGQGPAGSPFGQLPG	PAGPPGEAGKPGQGVPGDLGAPPSGPAGE----	413
DB	462	GPAPGAGERGQGPAGSPFGQLPG	PAGPPGEAGKPGQGVPGDLGAPPSGARGERP	521
QY	414	-----PGP-----	-----TGLPGPPGERGGSGRFFGADGVA	447
DB	522	GERGVQPPPGAPGPRGANGAP	CNDGAKGDGAPGAPGSGQAPGLGMP	581
QY	448	GERSGPPGAPGKSGPGE-----	AGRPEAGLFGAK-----GLTSGPSGSPGD	489
DB	582	GDRGDAGPKADGSPGKDGVRGLTG	PIGPPGPAGAPDGKSGSPGAPGPTG	641
QY	490	GKTGPPGAGQDGRPPGPPGARGQ----	-----AGVMGFP	534
DB	642	GEPGPPGAGFAGPGADGQGP	AKGEPGDAGAKGDAGPPGAPGAPGPTGNV	701
QY	535	GERGVPPGPPGA-----	VGPAKDGGEAGAQGPPGPAGP-----	582
DB	702	GARGSAGPPGATGPPGAAGRVG	PPGPSNAGPPGPPGAKGEGKGRGETG	761
QY	583	GLPGAPGPPGAGKPGEGVPGDLG	APGSPGAPGPGTGLPPGPRGGRGSGRFFGAD	642
DB	762	GPPGPPGAGKSGFSGADGAPAG	PTGPTPGQIAGQGVVGLPGQGERGFGL	821
QY	643	GVAGPKGPAGSRGSPGAPG--	KGSPGEAGRPGEAGLPGAKGLTGS	699
DB	822	GKQPSGNSGERPPGPMGPPGLAG	PPGSPGREGAPGAGSGPRDGS	881
QY	700	GPAQDGRPGPPPGARGQAGVMG	FPKGAAGEPGKAGERGVPPGAVGPAGKDEG--	758
DB	882	GPPGAPGAPGAPGVGVPAGKSG	DRGETGPAGPAGVPVGPAGARGPAGQ	941
QY	759	-----AGAGGPPG----	-----AGPAGERGQGPAGSP-----	792
DB	942	GDRGIKHRGFGSLQGP	PPPGSGPGEQGPAGPRPPGSGAGAPK	1001
QY	793	GPPGEAGKPGQGVPGDLGAP	GPSPGAPG	820
DB	1002	GPPGPRGRTGDAGVPVPPGPPG	PPGPPG	1029

RESULT 5

RESULT 3
 US-08-963-825-18
 ; Sequence 18, Application US/08963825
 ; Patent No. 6110689
 ; GENERAL INFORMATION:
 ; APPLICANT: Qvist, Per
 ; APPLICANT: Bonde, Martin
 ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 ; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
 ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
 ; TITLE OF INVENTION: Disorders Associated with the Metabolism of
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby PC
 ; STREET: 805 Third Avenue

Db 601 GLUGMPGEGGAAGLPGFKGDRGDAGPKGADGSPGKDGVRGLTGPITGPPGPAAGPDKGES 660
Qy 478 -----GLTSGPCSPDPDKTTPPPAGODGRPPGPPGARGO-----A 516
Db 661 GPSGPAPTGARGAPCDRGEPGPPGAPGPPGADGQFGAKGEPGDAGAKGDAGPPCPA 720
Qy 517 GWNVFFPGKGAAGEPKAGRGVVPBPAA-----VGPAGKDGEGAAGQPPGAPG- 566
Db 721 GPAGPPGPICNVGAPCAKAGARGAGSPPGATGTPFGAAGRVGPPGPGSCNAGPPGPPGAPKE 780
Qy 567 --AGERGEOPGAGSPFOGLPGPAGPPGAEAGKPGGQVPGDLGAPGSPGAGEPCPTGLP 624
Db 781 GKGPRGETGPAGRGPEVPPPPGPPGAGSKGSPGADGAPAGTGPPOGIAQOGVWGLP 840
Qy 625 GPPEERGSGPSRFFPGADGVAGPKPAGRGSPGPAAGP---KGSPEAGRGPEAGLPGAK 681
Db 841 GQGERGFFGLPGPSEPKQKQPSGASGERGPPGPMGPPGLAGPGEESGREGAPGAEGSP 900
Qy 682 GLTSGPCSPDPDKTTPPPGAPQDGRPPGPPGARGQAGVWVFFGPKGAAGEPKAGER 741
Db 901 GRDGSPGAKGDRGETGPAGPPGAXGAXGAPGVPGPAGKSGDRGETGPAGPAGVPAGAR 960
Qy 742 GVDGPPGAVGPAKGDE-----AGAQGPPG-----AGPAGERGEQ 777
Db 961 GPAGPQPRGDRKETGEQDGRGIKHHRGFSGLQPPGPPGSGPEQGPSGASGAPGRGPP 1020
Qy 778 GPAGSP---GFOGLPGPAGPPGAEAGKPGGQVPGDLGAPRGSPGAP 820
Db 1021 GSAGAPKQGLNGLPGIPIGPPRGRTGDAGVPVGGPPGPPGPPG 1066

RESULT 6
US-09-500-811-18
; Sequence 18, Application US/09500811
; Patent No. 632314
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:

QY	643	GVAGPKPAGERGSPGAGPKGSPGAGRPGAEAGLPGA	-----KGLT	684
Db	824	GOQAPGASGDRGPPGPPVGGTGPAGFGREGSPGADGPPGRDGAAGVKGDRCGETGAV	883	
QY	685	GSPGSPGDDKTKTGGPGP	-----AGODGRGPPGPGAGRGQAGVGMGFPKGAAGEP	735
Db	884	GAFCAFPSPGPGAGFTKQGDGRGAGAGQPMGSPGAGARGIQGPQGPGRDGKEAGEP	943	
QY	736	GK---AGERGVPGPAGVAGPKGDKGAGAGQGPAGPAGERGEQGPAGSPGFGQLPGA	792	
Db	944	GERGLKHRRFTGLQLPGPPGSPGQAGSPAGSPGPRGPPGVPVPSGDKGANGIPGI	1003	
QY	793	GPPEGKGEQGVPGDLGAPGSPGAG	820	
Db	1004	GPPGPRGRSGETGPAGPPGNPGPPPGP	1031	
RESULT 13				
US-08-963-825-20				
; Sequence 20, Application US/08963825				
; Patent No. 6110689				
; GENERAL INFORMATION:				
; APPLICANT: Qvist, Per				
; APPLICANT: Bonde, Martin				
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments				
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the				
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of				
; TITLE OF INVENTION: Disorders Associated with the Metabolism of				
; NUMBER OF SEQUENCES: 21				
; CORRESPONDENCE ADDRESS:				
; ADDRESSEE: Darby & Darby PC				
; STREET: 805 Third Avenue				
; CITY: New York				
; STATE: New York				
; COUNTRY: USA				
; ZIP: 10022				
; COMPUTER READABLE FORM:				
; MEDIUM TYPE: Floppy disk				
; COMPUTER: IBM PC compatible				
; OPERATING SYSTEM: PC-DOS/MS-DOS				
; SOFTWARE: PatentIn Release #1.0, Version #1.25				
; CURRENT APPLICATION DATA:				
; APPLICATION NUMBER: US/08/963,825				
; FILING DATE:				
; CLASSIFICATION: 436				
; PRIOR APPLICATION DATA:				
; APPLICATION NUMBER: US/08/187,319				
; FILING DATE: 21-JAN-1994				
; ATTORNEY/AGENT INFORMATION:				
; NAME: Gogoris, Adda C				
; REGISTRATION NUMBER: 29,714				
; REFERENCE/DOCKET NUMBER: 4305/08701				
; TELECOMMUNICATION INFORMATION:				
; TELEPHONE: 212-527-7700				
; TELEFAX: 212-753-6237				
; TELEX: 236687				
; INFORMATION FOR SEQ ID NO: 20:				
; SEQUENCE CHARACTERISTICS:				
; LENGTH: 1418 amino acids				
; TYPE: amino acid				
; TOPOLOGY: linear				
; MOLECULE TYPE: protein				
; ORIGINAL SOURCE:				
; ORGANISM: Homo sapiens				
; IMMEDIATE SOURCE:				
; CLONE: COLLAGEN -ALPHA 1 (II)				
US-08-963-825-20				
Query Match 53.4%; Score 2476; DB 3; Length 1418;				
Best Local Similarity 50.9%; Pred. No. 1.4e-152;				
Matches 503; Conservative 45; Mismatches 272; Indels 168; Gaps 17;				
QY	1	GPPEPDPDTGLPDPGRGPGFPGADGVAGPKPAGERGSPGFPAGPKGSPGAEGRP	60	

```
Db 156 GNPGEFGVSGMPGPPGPPGPKGDDGAGKPKGAGRGPPGPGAGGFFGTTGGLP 215
Qy 61 -----GEAGLPKAGLGTSP---GSPGP-----DKTGTGPPGA--- 90
Db 216 GVKGHRGVPGLDGAKGAGAGVKGESGSGSPGPMGRGLPCRGRGTGPAGAAGAR 275
Qy 91 GQDGRPPGPPGARGQGVGMGFP-----GPKGAAGFPKAGRGVPPGPGAVGPA 141
Db 276 GNDGQPGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 335
Qy 142 -----GKDGAGAGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 195
Db 336 GASGNPDTGTPGAKGSAGAGIAGAPFPGRGPPDPQATGTPGKGTGKPIAGFK 395
Qy 196 GD---LGAPSGPAGFPPTGLP-----GPPGERGSGSRGFFPGADGVA 237
Db 396 GEQPKGFPAGPQAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 455
Qy 238 GPKGAPAGRGSPGAPGKSGPAGRGSPGAPGKSGPAGRGSPGAPGKSGPAGRG 297
Db 456 GPKGAPAGRGSPGAPGKSGPAGRGSPGAPGKSGPAGRGSPGAPGKSGPAGRG 515
Qy 298 GRGPPGPPGARGQGVGMGFPKCAAGFPKAGRGVPPGPPGAVGPPAGKDGAGAG 357
Db 516 GRGPPGPPGARGQGVGMGFPKCAAGFPKAGRGVPPGPPGAVGPPAGKDGAGAG 575
Qy 358 GPAGPAGRGSPGAPGKSGPAGRGSPGAPGKSGPAGRGSPGAPGKSGPAGRG 413
Db 576 GPAGPAGRGSPGAPGKSGPAGRGSPGAPGKSGPAGRGSPGAPGKSGPAGRG 635
Qy 414 -----PGTGLP-----GPPGERGSGSRGFFPGADGVPAGPKPA 447
Db 636 GERGSFGAQLGQGRGLPTGTGTPGKASGAPGPPGAGQPPGLQMPGGERGAG 695
Qy 448 GERGS-----PGPAGPKSGPAGRGSPGAPGKSGPAGRGSPGAPGKSGP 483
Db 696 GURGVGKPEGAPKDGGRGLTGPPIPPGPPGANGKEGVPPEPPGAGSAGARGP 755
Qy 484 GSPGDPGTGTPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 534
Db 756 GGTGPPGTSIAGPPGADGPPGKAGGPPGPPGPPGPPGPPGPPGPPGPPGPP 815
Qy 535 GERGVGPPGA-----VGPAGKAGAGAGPPGPA---GPAGRGSPGAPGPPG 582
Db 816 GARGAQGPPGATGPPGAGRVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 875
Qy 583 GLPGAPGPPGKAGKPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 642
Db 876 GLQGPAGPPGKPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 935
Qy 643 GVAGPKPAGRGSPGAPGKSGPAGRGSPGAPGKSGPAGRGSPGAPGKSGPAG 684
Db 936 GQGGAPGASGDKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 995
Qy 685 GSPGSPGDKTGTGPP-----AGQDGRPPGPPGPPGPPGPPGPPGPPGPP 735
Db 996 GAFGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1055
Qy 736 GK---AGRGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 792
Db 1056 GERGLKHRGFTGLQGLFPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1115
Qy 793 GPPGAGKPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 820
Db 1116 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1143
```

RESULT 14

US-09-010-999-1

; Sequence 1, Application US/09010999

; Patent No. 6132976

; GENERAL INFORMATION:

```
; APPLICANT: Poole, Anthony R.
; APPLICANT: Hollander, Anthony P.
; APPLICANT: Billinghamurst, R. C.
; TITLE OF INVENTION: IMMUNOASSAYS FOR THE MEASUREMENT OF
; TITLE OF INVENTION: COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,999
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 4335
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/448,501
; FILING DATE: 17-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/984,123
; FILING DATE: 04-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 032931/0212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Human Type II Collagen
; US-09-010-999-1
```

Query Match 53.4%; Score 2476; DB 3; Length 1418;

Best Local Similarity 50.9%; Pred. No. 1.4e-152;

Matches 503; Conservative 45; Mismatches 272; Indels 168; Gaps 17;

```
Qy 1 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 60
Db 156 GNPGEFGVSGMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 215
Qy 61 -----GEAGLPKAGLGTSP---GSPGP-----DKTGTGPPGA--- 90
Db 216 GVKGHRGVPGLDGAKGAGAGVKGESGSGSPGPMGRGLPCRGRGTGPAGAAGAR 275
Qy 91 GQDGRPPGPPGARGQGVGMGFP-----GPKGAAGFPKAGRGVPPGPGAVGPA 141
Db 276 GNDGQPGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 335
Qy 142 -----GKDGAGAGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 195
Db 336 GASGNPDTGTPGAKGSAGAGIAGAPFPGRGPPDPQATGTPGKGTGKPIAGFK 395
Qy 196 GD---LGAPSGPAGFPPTGLP-----GPPGERGSGSRGFFPGADGVA 237
Db 396 GEQPKGFPAGPQAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 455
Qy 238 GPKGAPAGRGSPGAPGKSGPAGRGSPGAPGKSGPAGRGSPGAPGKSGPAGRG 297
```


Qy	535	GERGVGPPGA-----VGPAGKDGAGAGQPPGA---GPAGERGEQGPAGSPGFQ	582
Db	816	GARGAQPPGATGFPGAAGRVGPPGSGNPNPFPFPGSGKDGKAGDSGPPGRAGEP	875
Qy	593	GLPGPAGPPGSAKPGEQGVPCDLGAPGSPGAPGPGTGLPGPPGERGPGSRGFPQAD	642
Db	876	GLQGPAGPPGKGPBGDDGSPGAEPPGFGQGLAGQGI VGLPGQGERGFPGLPGPSGEP	935
Qy	643	GVAGPKGPAGERSGSPGAPGPKGSPGEGAGRPGEAGLPGA-----KGLT	684
Db	936	GOQAGAPGASGDRGPPGVPFGLTGAPGEPGREGSPGADGPPGRDGAAGVKGDRGETGAV	995
Qy	685	GSPGSPGPDGKTGPBP-----AGQDRGPPGPPGARGOAGVMGFPCKGAAGEP	735
Db	996	CAPGAPGPPGSPGAPGTPGKQDRGEAGAGQPMGSPGAPAGIGQPGQPRGDKGEAGEP	1055
Qy	736	GK---AGERGVGPPGAVGPAGKDGAGAGQPPGAPGAGERGQGPAGSPGQGLPGPA	792
Db	1056	GERGLKGRGFTGLQGLPGPPGSDQAGSGPAGSPGPRGPPGVGPGSKDGANGIPGI	1115
Qy	793	GPPGAGKPGEGGVPCDLGAPGSPGAG	820
Db	1116	GPPGPRGRSGETGAPGPFNPGPPG	1143

Search completed: June 17, 2005, 15:17:00
 Job time : 41.8485 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 15:02:42 ; Search time 89.9315 Seconds
(without alignments)
3505.413 Million cell updates/sec

Title: US-10-658-989A-3

Perfect score: 4640

Sequence: 1 GPPGPGTGLPGRGGRGG.....GEQVPGDLGAPGSPGAGG 821

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1714042 seqs, 38379560 residues

Total number of hits satisfying chosen parameters: 1714042

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4640	100.0	821	17	US-10-658-989A-3
2	3482	75.0	617	17	US-10-658-989A-2
3	2769	59.7	1014	17	US-10-901-816A-6
4	2769	59.7	1014	17	US-10-901-816A-10
5	2767	59.6	1014	17	US-10-901-816A-9
6	2767	59.6	1014	17	US-10-901-816A-11
7	2765.5	59.6	1449	15	US-10-402-089-8
8	2765.5	59.6	1449	15	US-10-402-072A-8
9	2765	59.6	1014	17	US-10-901-816A-5
10	2765	59.6	1057	15	US-10-104-889-20
11	2765	59.6	1107	15	US-10-104-889-11
					Sequence 3, Appli
					Sequence 2, Appli
					Sequence 6, Appli
					Sequence 10, Appli
					Sequence 9, Appli
					Sequence 11, Appli
					Sequence 8, Appli
					Sequence 5, Appli
					Sequence 20, Appli
					Sequence 11, Appli

12	2765	59.6	1171	15	US-10-104-889-8	Sequence 8, Appli
13	2765	59.6	1388	15	US-10-104-889-10	Sequence 10, Appli
14	2765	59.6	1464	14	US-10-216-705-21	Sequence 21, Appli
15	2765	59.6	1464	16	US-10-357-851-1	Sequence 1, Appli
16	2765	59.6	1464	16	US-10-358-024-1	Sequence 1, Appli
17	2765	59.6	1464	16	US-10-788-792-150	Sequence 150, Appli
18	2761	59.5	1461	16	US-10-468-091-25	Sequence 25, Appli
19	2761	59.5	1464	15	US-10-291-265-243	Sequence 243, Appli
20	2759	59.5	1057	15	US-10-104-889-16	Sequence 16, Appli
21	2759	59.5	1014	17	US-10-901-816A-7	Sequence 7, Appli
22	2755	59.4	1464	10	US-09-918-715-261	Sequence 261, Appli
23	2755	59.4	1464	14	US-10-060-036-159	Sequence 159, Appli
24	2755	59.4	1464	14	US-10-171-311-36	Sequence 36, Appli
25	2755	59.4	1464	14	US-10-149-352-2	Sequence 2, Appli
26	2755	59.4	1464	14	US-10-177-293-65	Sequence 65, Appli
27	2755	59.4	1464	14	US-10-301-822-28	Sequence 28, Appli
28	2755	59.4	1464	16	US-10-734-564-79	Sequence 79, Appli
29	2755	59.4	1464	16	US-10-474-794-261	Sequence 261, Appli
30	2755	59.4	1464	16	US-10-723-860-2289	Sequence 2289, Appli
31	2755	59.4	1464	17	US-10-852-335A-157	Sequence 157, Appli
32	2754	59.4	1014	17	US-10-901-816A-8	Sequence 8, Appli
33	2752.5	59.3	1463	15	US-10-402-089-2	Sequence 2, Appli
34	2752.5	59.3	1463	15	US-10-402-072A-2	Sequence 2, Appli
35	2752	59.3	1014	17	US-10-901-816A-13	Sequence 13, Appli
36	2750	59.3	1169	15	US-10-104-889-6	Sequence 6, Appli
37	2732.5	58.9	1453	16	US-10-468-091-26	Sequence 26, Appli
38	2729	58.8	1014	17	US-10-901-816A-12	Sequence 12, Appli
39	2704	58.3	1341	14	US-10-058-124-18	Sequence 18, Appli
40	2511	54.1	822	15	US-10-342-331-49	Sequence 49, Appli
41	2488	53.6	1017	16	US-10-639-286-10	Sequence 10, Appli
42	2479	53.4	1014	14	US-10-194-441A-1	Sequence 1, Appli
43	2476	53.4	1418	14	US-10-058-124-20	Sequence 20, Appli
44	2476	53.4	1418	16	US-10-468-091-5	Sequence 5, Appli
45	2460	53.0	1014	14	US-10-194-441A-48	Sequence 48, Appli

ALIGNMENTS

RESULT 1

US-10-658-989A-3
; Sequence 3, Application US/10658989A
; Publication No. US20050101531A1
; GENERAL INFORMATION:
; APPLICANT: BOWSTRA, Jan Bastiaan
; APPLICANT: YUZO, Toda
; TITLE OF INVENTION: Use of recombinant gelatin-like proteins as plasma expanders and
; FILE REFERENCE: BOWSTRA-3
; CURRENT APPLICATION NUMBER: US/10/658,989A
; CURRENT FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: EP 02078745.3
; PRIOR FILING DATE: 2002-09-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 821
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Hu-4
US-10-658-989A-3

Query Match 100.0%; Score 4640; DB 17; Length 821;

Best Local Similarity 100.0%; Pred. No. 2.9e-225;

Matches 821; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPPGPGTGLPGRGGRGGPGAGVAGPKGPGAGPKGSGPAGPKGSGPAGP 60

Db 1 GPPGPGTGLPGRGGRGGPGAGVAGPKGPGAGPKGSGPAGPKGSGPAGP 60

Qy 61 GEAGLPGAKGLTGTSGPSGPGPGKGTGPPGACODGPPGPPGARGQGVGFGPKCAA 120

Db 61 GEAGLPGAKGLTSGSPGPGDKTGPAGPAGQDGRPGPPGARGQAGVMGFPKGA 120
 QY 121 GEPGKAGRGVPPGAVGAGKAGCAQCPGPPAGPAGRGQAGSPGFGQLPGPA 180
 Db 121 GEPGKAGRGVPPGAVGAGKAGCAQCPGPPAGPAGRGQAGSPGFGQLPGPA 180
 QY 181 GPPGAGKRGQGVPGDLGAPGSPGAGEPPTGLPGPPGRRGFGSPGADGAGPK 240
 Db 181 GPPGAGKRGQGVPGDLGAPGSPGAGEPPTGLPGPPGRRGFGSPGADGAGPK 240
 QY 241 GPAGRRGSPGAGPKGSPGAGRPCEAGLPGCAKGLTSGSPGPDGKTGPPGAGQDGR 300
 Db 241 GPAGRRGSPGAGPKGSPGAGRPCEAGLPGCAKGLTSGSPGPDGKTGPPGAGQDGR 300
 QY 301 GPPGPPGARGQAGVMGFPKGAAGRGKAGRGVPPGAVGAGKAGCAQCPGPA 360
 Db 301 GPPGPPGARGQAGVMGFPKGAAGRGKAGRGVPPGAVGAGKAGCAQCPGPA 360
 QY 361 GPAGRRGQAGSPGFGQLPGPAGPAGRGKAGRGVPPGAVGAGKAGCAQCPGPA 420
 Db 361 GPAGRRGQAGSPGFGQLPGPAGPAGRGKAGRGVPPGAVGAGKAGCAQCPGPA 420
 QY 421 GPPGRRGSPGRRGFGPAGDVGAGKAGRGVPPGAVGAGKAGCAQCPGPA 480
 Db 421 GPPGRRGSPGRRGFGPAGDVGAGKAGRGVPPGAVGAGKAGCAQCPGPA 480
 QY 481 GSPGSPGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGAAGRGKAGRGV 540
 Db 481 GSPGSPGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGAAGRGKAGRGV 540
 QY 541 GPPGAVGAGKAGCAQCPGPPGARGQAGVMGFPKGAAGRGKAGRGV 600
 Db 541 GPPGAVGAGKAGCAQCPGPPGARGQAGVMGFPKGAAGRGKAGRGV 600
 QY 601 GVPDGLGAPGSPGAGPFGTGLPGPPGRRGFGSPGADGAGPKAGLPGCAKGLT 660
 Db 601 GVPDGLGAPGSPGAGPFGTGLPGPPGRRGFGSPGADGAGPKAGLPGCAKGLT 660
 QY 661 GPKGSPGAGRPBAGLPGCAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGARGQ 720
 Db 661 GPKGSPGAGRPBAGLPGCAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGARGQ 720
 QY 721 GVMGFPKGAAGRGKAGRGVPPGAVGAGKAGCAQCPGPA 780
 Db 721 GVMGFPKGAAGRGKAGRGVPPGAVGAGKAGCAQCPGPA 780
 QY 781 GSPGFGQLPGPAGPAGRPCEAGKPGQGVPGDLGAPGSPGAGG 821
 Db 781 GSPGFGQLPGPAGPAGRPCEAGKPGQGVPGDLGAPGSPGAGG 821

RESULT 2
 US-10-658-989A-2
 ; Sequence 2, Application US/10658989A
 ; Publication No. US20050101531A1
 ; GENERAL INFORMATION:
 ; APPLICANT: BOWSTRA, Jan Bastiaan
 ; APPLICANT: YUZO, Toda
 ; TITLE OF INVENTION: Use of recombinant gelatin-like proteins as plasma expanders and
 ; FILE REFERENCE: BOWSTRA-3
 ; CURRENT APPLICATION NUMBER: US/10/658,989A
 ; PRIOR FILING DATE: 2003-09-10
 ; PRIOR APPLICATION NUMBER: EP 02078745.3
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 617
 ; TYPE: PRT
 ; ORGANISM: Artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: Hu-3

US-10-658-989A-2
 Query Match 75.0%; Score 3482; DB 17; Length 617;
 Best Local Similarity 100.0%; Pred. No. 2.5e-167;
 Matches 616; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GPPGRRGPTGLPGPPGRRGFGSPGADGAGPKAGLPGCAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGA 120
 Db 1 GPPGRRGPTGLPGPPGRRGFGSPGADGAGPKAGLPGCAKGLTSGSPGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGA 120
 QY 61 GEAGLPGAKGLTSGSPGSPGDKTGPAGPAGQDGRPGPPGARGQAGVMGFPKGA 120
 Db 61 GEAGLPGAKGLTSGSPGSPGDKTGPAGPAGQDGRPGPPGARGQAGVMGFPKGA 120
 QY 121 GEPGKAGRGVPPGAVGAGKAGCAQCPGPPAGPAGRGQAGSPGFGQLPGPA 180
 Db 121 GEPGKAGRGVPPGAVGAGKAGCAQCPGPPAGPAGRGQAGSPGFGQLPGPA 180
 QY 181 GPPGAGKRGQGVPGDLGAPGSPGAGEPPTGLPGPPGRRGFGSPGADGAGPK 240
 Db 181 GPPGAGKRGQGVPGDLGAPGSPGAGEPPTGLPGPPGRRGFGSPGADGAGPK 240
 QY 241 GPAGRRGSPGAGPKGSPGAGRPCEAGLPGCAKGLTSGSPGPDGKTGPPGAGQDGR 300
 Db 241 GPAGRRGSPGAGPKGSPGAGRPCEAGLPGCAKGLTSGSPGPDGKTGPPGAGQDGR 300
 QY 301 GPPGPPGARGQAGVMGFPKGAAGRGKAGRGVPPGAVGAGKAGCAQCPGPA 360
 Db 301 GPPGPPGARGQAGVMGFPKGAAGRGKAGRGVPPGAVGAGKAGCAQCPGPA 360
 QY 361 GPAGRRGQAGSPGFGQLPGPAGPAGRGKAGRGVPPGAVGAGKAGCAQCPGPA 420
 Db 361 GPAGRRGQAGSPGFGQLPGPAGPAGRGKAGRGVPPGAVGAGKAGCAQCPGPA 420
 QY 421 GPPGRRGSPGRRGFGPAGDVGAGKAGRGVPPGAVGAGKAGCAQCPGPA 480
 Db 421 GPPGRRGSPGRRGFGPAGDVGAGKAGRGVPPGAVGAGKAGCAQCPGPA 480
 QY 481 GSPGSPGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGAAGRGKAGRGV 540
 Db 481 GSPGSPGPDGKTGPPGAGQDGRPGPPGARGQAGVMGFPKGAAGRGKAGRGV 540
 QY 541 GPPGAVGAGKAGCAQCPGPPGARGQAGVMGFPKGAAGRGKAGRGV 600
 Db 541 GPPGAVGAGKAGCAQCPGPPGARGQAGVMGFPKGAAGRGKAGRGV 600
 QY 601 GVPDGLGAPGSPGAG 616
 Db 601 GVPDGLGAPGSPGAG 616

RESULT 3
 US-10-901-816A-6
 ; Sequence 6, Application US/10901816A
 ; Publication No. US20050058703A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Chang, Robert C.
 ; APPLICANT: Olsen, David R.
 ; APPLICANT: James, Polarek W.
 ; APPLICANT: Williams, Kim E.
 ; TITLE OF INVENTION: Gelatin Capsules
 ; FILE REFERENCE: FP0404 US
 ; CURRENT APPLICATION NUMBER: US/10/901,816A
 ; CURRENT FILING DATE: 2004-07-29
 ; PRIOR APPLICATION NUMBER: US 60/492,085
 ; PRIOR FILING DATE: 2003-08-01
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 6
 ; LENGTH: 1014
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; OTHER INFORMATION: Hu-3


```

Db      672  PGTGPPGAGKAGGAGGATGTFGAAGRVGPPFPGSGNAGPPGPPGACKEGKGP 731
QY      570  RGEQGPAGSPGQGLGPPAGPPGAGKPGEGQGVGDLGAPGSPGAGEPPTGLPGPPGE 629
Db      732  RGETGPAGRPGVEGPPGPPGAGEKSGFADGAPAGTGTGPGQIAGQGVVGLPGQGE 791
QY      630  RGGPGSRGPPGADGVAGPKGAPGAGRGSGPPAGP---KSGPGEAGRPGEAGLPGAKGLTGS 686
Db      792  RGFTGLPFGSPGPKQKQSGASGERGPPGPMGPPGLAGPPGEGSREGAPCAEGSPGRDGS 851
QY      687  PGSPGPDGKTGPPGAGQDGRPPGPPGARGQAGVGMFPGPKGAAGEPKKAGRGVPPG 746
Db      852  PKAGKDRGETGPAGPPGAPGAPGVPVGGPAGKSGDRGETGTPAGPAGVPVGPAGARGPAG 911
QY      747  PGAVGPAGKQGE-----AGAQGPPGP-----AGPAGGERGQGPAGS 782
Db      912  QGPRGDKGTGEGQDGRGKKGHRGFSGLQGGPPGSGEGQGPAGPRGPPGSAGA 971
QY      783  P---GFOGLPGPAGPPGAEAGKPGQGVGDLGAPGSPGAG 820
Db      972  PGKDLNGLPGPIGPPGPRGRTGDAGVPVGGPPGPPGPPG 1012

```

RESULT 5

```

US-10-901-816A-9
; Sequence 9, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; PRIOR FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1014
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-901-816A-9

```

Query Match 59.6%; Score 2767; DB 17; Length 1014;
 Best Local Similarity 55.3%; Pred. No. 2.4e-131;
 Matches 554; Conservative 34; Mismatches 231; Indels 182; Gaps 16;

```

QY      1  GPPGEGPTGLPDPGGERGSGRGP-----GADGVAGPKPAGERGSGP 47
Db      13  GPPGAPGQPGQFQPPGEPGPGASG-PMGPPPPGPPGKNGDDGEAGKRGPRGERGPPG 71
QY      48  AGPKSGFAGRPGEGAGLCAKGLTSGSPGPDGKTGPPGAGQDGRPPGPPGARGQ 107
Db      72  QGARGLFTAGLPMKKGHRGFSGLDAGKADGAPGKSGPSPGNGAPQMGPPPLFGE 131
QY      108  AGVMGPPGKGAAGEFGKAGRGVPPGAVGP-----AGKDGAGAGQPPGPPAGPAGE 161
Db      132  RGRPGAPGAGARGNDGATGAAGPPGPTGPPGPPGAVGAKGAGGQGRGSGPGGV 191
QY      162  RGEQGP-----AGSPGQGLPGPAGPPGAEAGKPG 191
Db      192  RGEPPGPGAGAGPAGNPGADGQPKAGANGAPGIAGAPGPPGARGSPGQGGPPGPP 251
QY      192  QGVPGDLGAP-----GPSGA-----GPPGPTGLPGPPGREGG 224
Db      252  KNSGSPGAPGSKDGTAGKGPVGVQPPGPPAGEBKRGARGEGTGLPGPPGERGG 311
QY      225  PGRGPPGADGVAGPKPAGERGSGPAGPKSGPGEAGRPGEAGLPGAKGLTSGSPGSGP 284

```

```

Db      312  PGRGPPGADGVAGPKPAGERGSGPAGPKSGPGEAGRPGEAGLPGAKGLTSGSPGSP 371
QY      285  DGTGTPPGAGQDGRPPGPPGARGQAGVGMFPPGPKGAAGPPGKAGRGVPPGPAVGP 344
Db      372  DGTGTPPGAGQDGRPPGPPGARGQAGVGMFPPGPKGAAGPPGKAGRGVPPGPAVGP 431
QY      345  AGKDGAGAGQAGPPGPPGAGGERGEGPAGSPGQGLPGPAGPPGAGKPGEGVPGDLGA 404
Db      432  AGKDGAGAGQAGPPGPPGAGGERGEGPAGSPGQGLPGPAGPPGAGKPGEGVPGDLGA 491
QY      405  PGPSPGAGE-----PGP-----TGLPGPPGPPGPPGSGRGP 434
Db      492  PGPSPGARGERPPGGERGVQGPAGPRGANGAPGNDGAKGDAGAPGAGSGQAGFLQGM 551
QY      435  PGADGVAGPKPAGGERGSGPAGPKSGPGE-----AGRPGEAGLPGA 476
Db      552  PGERGAAGLPGPKGDRGADGPKGADGSGFKDGVRLTGTPIGPPGAGAPGPPGSGSPSG 611
QY      477  KGLTSGSPGPPGDKGTGTPPGPPAGQDGRPPGPPGARGQ-----AGVMGF 521
Db      612  AGPTGARGAPGDRGEPGPPGAGFAGPAGADGQPKAGKPGDAGKAGDAGPPGAPGAPG 671
QY      522  PGPKAAGEPKKAGRGVPPGPPGA-----VGPAGKDGAGAGQGGPPGAPG---AGE 569
Db      672  PGP1GPPGAPGAKGARGSAGPPGATGFPGAAGRVGPPGPPGPPGPPGAGKEGKGP 731
QY      570  RGEQGPAGSPGQGLGPPAGPPGAGKPGEGQGVGDLGAPGSPGAGEPPTGLPGPPGE 629
Db      732  RGETGPAGRPGVEGPPGPPGAGEKSGFADGAPAGTGTGPGQIAGQGVVGLPGQGE 791
QY      630  RGGPGSRGPPGADGVAGPKPAGERGSGPPGAPG---KSGPGEAGRPGEAGLPGAKGLTGS 686
Db      792  RGFTGLPFGSPGPKQKQSGASGERGPPGPMGPPGLAGPPGEGSREGAPCAEGSPGRDGS 851
QY      687  PGSPGPDGKTGPPGAGQDGRPPGPPGARGQAGVGMFPPGPKGAAGEPKKAGRGVPPG 746
Db      852  PKAGKDRGETGPAGPPGAPGAPGVPVGGPAGKSGDRGETGTPAGPAGVPVGPAGARGPAG 911
QY      747  PGAVGPAGKQGE-----AGAQGPPGP-----AGPAGGERGQGPAGS 782
Db      912  QGPRGDKGTGEGQDGRGKKGHRGFSGLQGGPPGSGEGQGPAGPRGPPGSAGA 971
QY      783  P---GFOGLPGPAGPPGAEAGKPGQGVGDLGAPGSPGAG 820
Db      972  PGKDLNGLPGPIGPPGPRGRTGDAGVPVGGPPGPPGPPG 1012

```

RESULT 6

```

US-10-901-816A-11
; Sequence 11, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; PRIOR FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1014
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-901-816A-11

```

Query Match 59.6%; Score 2767; DB 17; Length 1014;
 Best Local Similarity 55.3%; Pred. No. 2.4e-131;
 Matches 554; Conservative 34; Mismatches 231; Indels 182; Gaps 16;

Qy	1	GPPPEPGPTGLPPPPGERGGPGSGRGP	-----GADGVAGPKGPAGERGSGP	47	
Db	13	GPPCAPQPGQPPGPEGEFCASG-PMGPPCPGPPGKNGKDDGAGKPRGERGPPGP	71		
Qy	48	AGPKSGPGEAGRPEAGLPGAKGLTSGSGPCPDGKTPGCPAGODGPPGPPGPGARQ	107		
Db	72	QKAGLPTTAGLPGMKHHRGFSGLDGAKDAGPAGPKGPGSGPENGAGPQMGPPGLPGE	131		
Qy	108	AGVMGFPKPKAAAGEPGKAGERGVGPPGAVGP	-----AGKDGAGAGQPPGPAGPAGE	161	
Db	132	RGRFGAPGAGNDGATGAAGPPGTPAGPGFPGAVKAGGAGQGPGRSGPGV	191		
Qy	162	RGEQGP	-----AGSPGQGLPGPAGPPGEAGKPGEE	191	
Db	192	RGEPPGPGAGAGPAGNPGADGQPKAKGANGAPGIAGAPGPPGARGSPGQPGGPPGP	251		
Qy	192	QGVPGDLGAP	-----GPSGA	-----GEPGPTGLPGPPGERGG	224
Db	252	KNSGEPGAPGSKDGTAKGFPVGVGPPPGPAGGEGKRGARGGEPGPTGLPGPPGERGG	311		
Qy	225	PGSRGFPDAGDVAGPKGPAGBERGSPGAPGPKGSPGEAGRPGEAGLPGAKGLTSGPGSGP	284		
Db	312	PGSRGFPDAGDVAGPKGPAGBERGSPGAPGPKGSPGEAGRPGEAGLPGAKGLTSGPGSGP	371		
Qy	285	DGKTGPPGAPGODGRPPGPPGARGQAGVMGFPKPKGAAAGSPGKAGRGVPPGPGAVGP	344		
Db	372	DGKTGPPGAPGODGRPPGPPGARGQAGVMGFPKPKGAAAGSPGKAGRGVPPGPGAVGP	431		
Qy	345	AKDGEAGAQQPPGAPGAPGERGEOGAGSPGFQGLPGPAGPPGKAGKPGEOGVPGDLGA	404		
Db	432	AKDGEAGAQQPPGAPGAPGERGEOGAGSPGFQGLPGPAGPPGKAGKPGEOGVPGDLGA	491		
Qy	405	PGPSGPAGE	-----PGP	-----TGLPGPPGERGGPGSGRGP	434
Db	492	PGPSGARGERFPGERGVQGPFGPAGPRGANGAPGNDGAKGDAGAPGSGOGAPGLQGM	551		
Qy	435	PGADGVAGPKGPAGBERGSPGAPGPKGSPGE	-----AGRPGEAGLPGA	476	
Db	552	PGERGAAGLPKPGKDRGDAGPKGADGSPKDGVRGLTGP1GPPGPAGAPGPPGSGSPSGP	611		
Qy	477	KGLTSGPSGSPGDKTGPBPAGQDGRPGPPGPGARQ	-----AGVMGF	521	
Db	612	AGPTGARGAPGDRGEPGPPGAGFAGPPGAGDQGPAGKGPEDGAKGDPGAGPPGAPGP	671		
Qy	522	PGPKGAAGEPKKAGRGVPGPPGA	-----VGPAGKDGAGAGQPPGPGAPG	569	
Db	672	PGP1GNVGAPKAKHARGAGPPGATGFGAAGRVPVPPGSGNAGPPPPGPKAGKGGKGP	731		
Qy	570	RGEQGPAGSPFQGLPGPAGPPGEAGKPGEOGVPGDLGAPGSGPAGBPGPTGLPGPPGE	629		
Db	732	RGETGPAGRPGEVGPVPPGPPGAGEKSGPDAGDPAGAPGTPPGQGLAGRGVVGLPGQGE	791		
Qy	630	RGGPGSRGFPDAGDVAGPKGPAGERGSPGAPG	-----KGSPPGABRPGGEAGLPGAKGLTGS	686	
Db	792	RGFFGLPGPSPGEPKQGPSGASGRGPPGMPGPPGLAGPPGSSGREGAPGASGSGPGRDGS	851		
Qy	687	PGSPGPDGKTGPPGAGODGRPPGPPGAPGAGQAGVMGFPGPKGAAGEPGKAGRGVGP	746		
Db	852	PGAKGDRGETGAPGPPGAPGAPGAPGPPVGPAGKSGDRGETGAPGAPGPPGAPGAPGAP	911		
Qy	747	PGAVGPAGKDE	-----AGAOGPPG	-----AGPAGERGEOGPAGS	782
Db	912	QGPBGKXGTEQGDRIK1KHGRFSGLOQFPFGPPGSGPGEQGPSGASGAPGPRGPGSAGA	971		
Qy	783	P---GFGQLPGPAGPPGKAGKPGEOGVPGDLGAPGPGSPAG	820		
Db	972	PGKDLGLNGLPG1GPPGPRGRTGDAVGPVPPGPPGPPGPPG	1012		

RESULT 7

US-10-402-089-8

; Sequence 8, Application US/10402089

```

; Publication No. US20040005663A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
;
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
;
; FILE REFERENCE: FPO402.3 CON
;
; CURRENT APPLICATION NUMBER: US/10/402,089
;
; CURRENT FILING DATE: 2003-03-26
;
; PRIOR APPLICATION NUMBER: US 09/709,700
;
; PRIOR FILING DATE: 2000-11-10
;
; NUMBER OF SEQ ID NOS: 72
;
; SOFTWARE: PatentIn version 3.2
;
; SEQ ID NO 8
;
; LENGTH: 1449
;
; TYPE: PRT
;
; ORGANISM: Sus scrofa
;
; US-10-402-089-8

```

Query Match	59.6%	Score 2765.5	DB 15	Length 1449	
Best Local Similarity	56.7%	Pred. No. 3.7e-131			
Matches	559	Conservative 32	Mismatches 226	Indels 169	
Qy	1	GPFGERTGLPGPPGERGGPGS-----RGPPGA-----	DGVAGPKPAGERGSPCA	48	
Db	193	GPTCAPGQFQPPGPGSPGBGASGMPGPPGPKNGKNGDDGAGKPRGPRGPPGPPQ		252	
Qy	49	GPXGSPGEAGRP-----GEAGLPCAKGLTGSPPSGPDPDKTGPPOGQD	93		
Db	253	GARGLPDTAGLPGMKHURGSGLDCAKGDAGPAGKGPSPGNGAPQMGPRGLPGR	312		
Qy	94	GRPGPPPPGARGO---AGVMGFPPGXGAAGE---PGKAGERGVPFPVAGVGPAGKDGEA	147		
Db	313	GRPGPPGARGNDGATGAAGPGPTGPAGPPGPGAVGAKGAGPQARGSESPQVGR	372		
Qy	148	GAGCPPGAPGAGERGGPP-----AGSPGQGLPGPAGPPGEEAGKPBEG	192		
Db	373	GEPPGPGAGAGPAGNPGADGQPGCKGANGAPGIAGPPGARGPSGQPSGPPGPK	432		
Qy	193	GVPGDLGAP-----GPSGA-----	GERPPTGLPGPPGERGPP	225	
Db	433	GNSEPPAGPSKGDGTAKGPGPTGTGVOGPPGAGEEGKRGARGPFPAGLPGPPGERGPP	492		
Qy	226	GSRGFPAGDGVAGPKGPAGRGSPGAPAGKXGCEAGRPGGAGLPKAGLGTGSPGSPGD	285		
Db	493	GSRGFPAGDGVAGPKGPAGRGSPGAPAGKXGCEAGRPGGAGLPKAGLGTGSPGSPGD	552		
Qy	286	GKTGPPGAPGQDRGPPGPPGARGQAGVMGFPPKGAAGPFGKAGERGVPFPAGVGA	345		
Db	553	GKTGPPGAPQDGRGPPGPPGARGQAGVMGFPPKGAAGPFGKAGERGVPFPAGVGA	612		
Qy	346	GKDGEAGAQPPGAPGAGERGEOGAPGSPGQGLPGPAGPPGAGKPGEOGVPDGLGAP	405		
Db	613	GKDGEAGAQPPGAPGAGERGEOGAPGSPGQGLPGPAGPPGAGKPGEOGVPDGLGAP	672		
Qy	406	GPSGPAGE-----PGP-----	TGLPDPGERGGPSGRFP	435	
Db	673	GPSGARGERFPGERGVQPPGPPGAPRGANGAPGNDGAKGDAGAPGAPGSGQAPGLQMP	732		
Qy	436	GADGVAGKPAAGERGSPGAPKXGSGE-----	AGRPGEAGLPGAAGLGTSPGSP	486	
Db	733	GERGAALGPFKGDGAGPKGADGAPGKDGVRGLTGP	IGPPGPAGAPGDRGETGFSGPA	792	
Qy	487	GPDKTGPAPGQDGRPPG---PGPPGARGQAGVMGF---	PPKGAAGRPKGKAGERGV	539	
Db	793	GPTGARGAPGDRGEPPGPPGAPGAPGADGQPKAKGGTGPDPGPIGSGVAPGPKKARGS	852		
Qy	540	PGPPGA-----	VGPAGKDGAGAGQPPGPPAGPAGE---	RGEQSGPAGSPGFOGLRPP	587
Db	853	AGPPGATGFPGAAGRVGPPGSPGNAGPPGPPGAPGAGKSGKPRGEGTGPAGRPGGAGP	912		
Qy	588	AGPPGEGAGKEQGVDPDLGAPGSPGAPGERPPTGLFPP	GPGERGGPSGRGFPFAGDGVAGP	647	

```

Db 913 PGAGKSGGADGAPAGATPGTQGLAGQGVVGLPGQGRGPPGLPGPSGSGPKQGP 972
Qy 648 KGPAGRGSGPGAPGKSGPCEAGRPAGLPGAKGLTSGP-----GSPGPD---GKTGPPGP 701
Db 973 SPSGSGRGPFGMPG---PGLAGPFGESGREGAPGABGSPGDRGAPGKDRGESGPAGP 1029
Qy 702 AQDGRPPGPPGPGARGQAGVMGFPFGPKGAAGEPKAGRGVPGFPGAVGAGKDG---758
Db 1030 PGAPGAPGAPGPGVPGPAGKSGDRGETGAPGAPGVPVPGARGPAGPQGRGKGTGEQGD 1089
Qy 759 -----AGAGPPGP-----AGPAGERGQGPAGSP-----GQGLPGPAGP 794
Db 1090 RGIKHGRGSLQGGPPGSGPCEQSGASGAPGPRGPGSAGAPKDKGLNGLPGTGP 1149
Qy 795 PGBAGKPGGQGVPGDLGAPGSPGAPG 820
Db 1150 PGRGRGTGDAGVPVPGPPGPPGPPG 1175

RESULT 8
US-10-402-072A-8
; Sequence 8, Application US/10402072A
; Publication No. US20040018592A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.2 CON
; CURRENT APPLICATION NUMBER: US/10/402,072A
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 8
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-402-072A-8

```

Query Match 59.6%; Score 2765.5; DB 15; Length 1449;
 Best Local Similarity 56.7%; Pred. No. 3.7e-131;
 Matches 559; Conservative 32; Mismatches 226; Indels 169; Gaps 21;

```

Qy 1 GPPGCEPPTGLPQPPGRRGPGS-----RGTFGA-----DGVAGPKPAGERGSPGPA 48
Db 193 GPPGAPGQFGPPGPFEGFASGPMGPRGPRGPKNGDDGEAGKPRGRRGPPGPQ 252
Qy 49 GPKSGPGEAGRP-----GEAGLPGAKGLTSGSPGPPGDKTGPAGQD 93
Db 253 GARGLPGTAGLPGMKHGRGSLDGAAGKAGPAGPKGEGSGENGAPQGMGFRGLPGER 312
Qy 94 GRPQPPGPPCARQ---AGVMGPPGPKGAAGE---PGKAGERGVPPGPPCAVGPAGKCGEA 147
Db 313 GRPQPPGPPAGARGNDGATGAAGPPGPTGAPGPPGPAVGAKEAGPQARGSEGPQVR 372
Qy 148 GAQPPGPPGAPGAGERGEGP-----AGSPGQGLPGPAGPPGAGERGKGEQ 192
Db 373 GEPGPPGAPAGAPNCPADQGPCKGANGAPGATAGAPGPGARGSPGQSPGSPGPK 432
Qy 193 GVPDGLGAP-----GPGSPA-----GEPGTPGLPFPGERGGP 225
Db 433 GNSGEPGAPGSKGDTCAKGEPGTGVQPPGPGAGEGKRGARGEPGAGLGGPGERGGP 492
Qy 226 GSRGFPAGDVGAPKGPAGERGSPGAPKSGPCEAGRPAGLPGAKGLTSGSPGSPGD 285
Db 493 GSRGFPAGDVGAPKGPAGERGSPGAPKSGPCEAGRPAGLPGAKGLTSGSPGSPGD 552
Qy 286 GKTGPPGAPGQDRGPPGPPGARGQAGVMGFPFGKAAAGPPKAGRGVPGPPGAVGPA 345

```

```

Db 553 GKTGPPGAPGQDRGPPGPPGARGQAGVMGFPFGPKGAAGEPKKAGERGVPPGPAVGA 612
Qy 346 GKDEAGAGQPPGPPGAPGAGERGEGQGPAGSPGQGLPGPAGPPGEGAGKPGQGVPGDLAGP 405
Db 613 GKDEAGAGQPPGPPGAPGAGERGEGQGPAGSPGQGLPGPAGPPGEGAGKPGQGVPGDLAGP 672
Qy 406 GPGSGPAGE-----PGP-----TGLPGRGERGPGSRGPP 435
Db 673 GPGSGARGRGFPGERGVQGPFGPAGPRGANGAPGNDGAKGADAGAPGQSGAPGLQGMP 732
Qy 436 GADGVAGPPKGPAGERGSGPPGAPKSGPGE-----AGRPGEAGLPGAKGLTSGSPG 486
Db 733 GERGAAGLPGPKGDRGDAGPKGADGAPKDGVRGLTGPPIGPPGAPAGDKGTGSPGPA 792
Qy 487 GPDGKTGPPGAPGQDRGPP---PGPPCARGQAGVMGF---PGPKGAAGEPKKAGERGV 539
Db 793 GPTGARGAPGDRGEPGPPGAPGAPGADGQPGAKGKGTPPGPIGSGVAGPPGKARGS 852
Qy 540 PGPARGA-----VGPAGKDGAGAGQPPGAPGAGE---RGEQGPAGSPGFGGLPGP 587
Db 853 AGPPGATGFPGAAGRVGPPGPGNAGPPGPPGAPGKSGKGRGETGPAGRPGEAGPPG 912
Qy 588 AGPPGEGAGKPGEQGVPGDLGAPGSPGAPGPTGLPQPPGRRGPGSRGPPGADGVAGP 647
Db 913 PGPAGEKSGPAGDGPAGAGTTFGQGLAGQGVVGLPQGRGERGFPGLPSPSGEPGQGP 972
Qy 648 KGPAGERGSPGAPGKSGPGEAGRPGEAGLPGAKGLTSGP---GSPGPD---GKTGPPGP 701
Db 973 SPSGSGRGPFGMPG---PGLAGPFGESGREGAPGASGSPRDGAPGPKDRGESGPAGP 1029
Qy 702 AQDGRPPGPPGPPCARGQAGVMGFPFGPKGAAGEPKAGRGVPGPAGVGPAGKDG---758
Db 1030 PGAPGAPGAPGVPVPGAPKSGDRGETGAPGAPGVPVPGARGPAGPQGRGKGTGEQGD 1089
Qy 759 -----AGAGPPGP-----AGPAGERGEGQGPAGSP---GFGQLPGPAGP 794
Db 1090 RGIKHGRGSLQGGPPGPGSGPCEQSGASGAPGPRGPGSAGAPKDKGLNGLPGTGP 1149
Qy 795 PGBAGKPGGQGVPGDLGAPGSPGAPG 820
Db 1150 PGRGRGTGDAGVPVPGPPGPPGPPG 1175

```

RESULT 9

```

US-10-901-816A-5
; Sequence 5, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 5
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-5

```

Query Match 59.6%; Score 2765; DB 17; Length 1014;
 Best Local Similarity 55.4%; Pred. No. 3e-131;
 Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

```

Qy 1 GPPGFPPTGLPFPGERGPGSRGFPFGADGVAGPKPAGERGSPGAPGKSGPGEA---57
Db 1 GPMGSPGRGLPGPPGAPGQFGQPPGEPGCPGASGPMGPRGPPGPKNGDDGEGAKP 60

```



```
QY 58 GRPGEAGLP---GAKGLTSGSPGSP-----DGKTGPPAGQDGRPPGPPGAR 105
Db 61 GRPGERGPPGQAGRLPGTGLPMKMGHGFSLDGAKGAGPAGPKGSPGSPGNGAP 120
QY 106 QGAGVMGPPGPKGAAGEPGKAGER-----GVPPGPGAVGPA-----GKDGEGAGQ 150
Db 121 GQMGPRGLPGERGRFGAPGAGRGNDGATGAAGPPGTPGAGPPGPGAVGAKGAGPQ 180
QY 151 GPPGAPGAGRGEGP-----AGSPGPGGLPQPA 180
Db 181 GPRGSEGGVGRGPPGPPGAGAGPAGNPGADGQPGAKGANGAPGAGPFGPGARGPS 240
QY 181 GPPGEAGKPGQGVPGDLGAP-----GPSGA-----GEPGPT 213
Db 241 GPQGGGPPGPKGNSGPGAGSGKSDTGAKEGPGVGVQPPGAGEGKRGAGEPGPT 300
QY 214 GLPDPGGRGPGSGRFFGADGAVGPKGAGERSGPGAGPKGSPGAGRPGEAGLPGAK 273
Db 301 GLPDPGGRGPGSGRFFGADGAVGPKGAGERSGPGAGPKGSPGAGRPGEAGLPGAK 360
QY 274 GLTSGPSGPDGKTGTPGPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEPGKAGER 333
Db 361 GLTSGPSGPDGKTGTPGPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEPGKAGER 420
QY 334 GVPDPGAVGPKDGEAGAGQPPGPGAGPAGERGEGQAGSPGQGLPGPAGPPGEGAKP 393
Db 421 GVPDPGAVGPKDGEAGAGQPPGPGAGPAGERGEGQAGSPGQGLPGPAGPPGEGAKP 480
QY 394 GEOGVPGDLGAPGSPGAGE-----PGP-----TGLPDP 423
Db 481 GEOGVPGDLGAPGSPGAGEGFGPGERGVQPPGPPGARGANGAPGNDGAKGDAGAP 540
QY 424 GERGGPSRPGPDAGDVAGPKGAGERSGPGAGPKGSPG-----AGRPGEAGLP 474
Db 541 GSQAGPLQGMGGERGAGLAGLPKGRDGDAGPKGADSGKDGVRGLTGPPIGPPGAPAG 600
QY 475 GAK-----GLTSGPSGPDGKTGTPGPGAGQDGRPPGPPGARGQ-----515
Db 601 GDKGESGSPGAPGTGARGAFGRGEPGPPGAGPAGPPGADGQPGAKGEPGCDAGAKGDA 660
QY 516 -----AGVMGPPGPKGAAGEPGKAGERGVPPGPA-----VGPAGKDGEGAGQPP 561
Db 661 GPPGAPGAGPPGPIGNVGAECAGKARGSPGATGPPGAGRVGPPGSPGNAGPPGPP 720
QY 562 GPAGP---AGERGEGPAGSPGQGLPGPAGPPEAGKPGGQGVPGDLGAPGSPGAGEP 618
Db 721 GPAGKEGKGRGETGTPAGRGVEGPPGPPGAGEKSGPDAGPAGTTPGQGLAGQ 780
QY 619 GPTGLPDPGRRGPGSGRFFGADGAVGPKGAGERSGPPGAPG---KGSPGEAGRPGEA 675
Db 781 GVVGLPQGRGERGPPGLPSPGSEPGKQFSGASGERGPPGPMGLAGPPGSGREGAP 840
QY 676 GLPAGKGLTSGSPGPDGKTGTPGPGAGQDGRPPGPPGARGQAGVMGPPGPKGAGEP 735
Db 841 GAEGSPGDRGSPGAKGDRGETGTPAGPPGAPGAPGVPVPGAGSGDRGETGTPAGPAGV 900
QY 736 GKAGERGVPPGPPGAVGPAKDGGE-----AGAQQGPPG-----AGPA 771
Db 901 GPAGARGPAGPQGRGDKETGEGQDRIKGRHGFSLGQPPGPPGSPGEGQPSGAGPA 960
QY 772 GERGEGPAGSP---GFQGLPGPAGPGEAGKPGEGQGVPGDLGAPGSPGAP 820
Db 961 GPRGPPGSAGAPGKODGLNGLPGPIGPPGRKGTGDAGVGPVPPGPPGPPG 1012
```

RESULT 10

```
US-10-104-889-20
; Sequence 20, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
```

```
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-10-104-889-20
```

```
Query Match 59.6%; Score 2765; DB 15; Length 1057;
Best Local Similarity 55.4%; Pred. No. 3.1e-131;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;
```

```
QY 1 GPPGEPGTGLPDPGRRGPGSGRFGPCADGAVGPKGAGERSGPPGAGPKGSGPGEA---57
Db 18 GPMGSPGRLGPPGAPGPGQFGQPGCEGPGSGMGPGRGPPGPKNGDDGEGAKP 77
QY 58 GRPGEAGLP---GAKGLTSGSPGSP-----DGKTGPPAGQDGRPPGPPGAR 105
Db 78 GRPGERPFPQAGRLPGTGLPMKMGHGFSLDGAKGAGPAGPKGSPGSPGNGAP 137
QY 106 QGAGVMGPPGPKGAAGEPGKAGER-----GVPPGPGAVGPA-----GKDGEGAGQ 150
Db 138 GQMGPRGLPGERGRFGAPGAGRGNDGATGAAGPPGTPGAGPPGPGAVGAKGAGPQ 197
QY 151 GPPGAPGAGRGEGP-----AGSPGPGGLPQPA 180
Db 198 GPRGSEGGVGRGPPGPPGAGAGPAGNPGADGQPGAKGANGAPGAGPFGPGARGPS 257
QY 181 GPPGEAGKPGQGVPGDLGAP-----GPSGA-----GEPGPT 213
Db 258 GPQGGGPPGPKGNSGPGAGSGKSDTGAKEGPGVGVQPPGAGEGKRGAGEPGPT 317
QY 214 GLPDPGGRGPGSGRFFGADGAVGPKGAGERSGPGAGPKGSPGAGRPGEAGLPGAK 273
Db 318 GLPDPGGRGPGSGRFFGADGAVGPKGAGERSGPGAGPKGSPGAGRPGEAGLPGAK 377
QY 274 GLTSGPSGPDGKTGTPGPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEPGKAGER 333
Db 378 GLTSGPSGPDGKTGTPGPGAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEPGKAGER 437
QY 334 GVPDPGAVGPAKDGEGAGAGQPPGPGAGPAGERGEGQAGSPGQGLPGPAGPPGEGAKP 393
```

Db 438 GVPGPAGVAGKDGAGAGQPPGAPGAGERGEQAGSPGQGLPGPAGPGEAGKP 497
Qy 394 GEQVPGDLGAPSGPAGE-----PGP-----TGLPGPP 423
Db 498 GEQVPGDLGAPSGPAGE-----PGP-----TGLPGPP 423
Qy 424 GERGGPSRPPGADGVAGPKGAPGAGSGSPGAPGKSGPGE-----AGRPGAGLP 474
Db 558 GSQAGPLQMGGERGAAGLPGPKDRGADGPKGADSGFGKGVGRGLTGTPTGPPAGAP 617
Qy 475 GAK-----GLTSGPSGPDGKTGPPGAGQDGRPPGPPGARGQ-----515
Db 618 GDRGESPSGAPGPTGARGAPDRGPPGPPGAPGAGPAGGQDQPKAGEPCDAGAKGDA 677
Qy 516 -----AGVMGFPKPGKAAGEPKGAGERGVPPGPA-----VGPAGKDEAGAGQPP 561
Db 678 GPPGAPGAPGPPGPIGNVAPKAGKARGSGAGPPGATGPPGAAGRVGPPGSGNAGPPGP 737
Qy 562 GPAGP---AGEGEQAGSPGFGQLPGPAGPPGAGKPGGQGVPGDLGAPGSPGAPGE 618
Db 738 GPAGKGGKPRGEGTGPAGRPGEVPPGPPGAPGKSGFADGAPAGPTGPPQGIAGQR 797
Qy 619 GPTGLPGRGGRGSRFPFGADGVAGPKGAPGAGSGSPGAPG---KSGPGEAGRPGEA 675
Db 798 GVVGLPQGRGERGFGFLRPSGEPCKQKQSGASGERGPPGMPGLAGPPGSGREGAP 857
Qy 676 GLPGAKGLTSGPSGPDGKTGPPGAGQDGRPPGAGQAGVMGFPKGAAGEP 735
Db 858 GAGGSPGRDPSGAKGDRGETGAPGPPGAPGAPGVPVGPAGKSGDRGETGAPGAPV 917
Qy 736 KXAGERGVPPGAVGAPKADGE-----AGAGQPPG-----AGPA 771
Db 918 GPAGARGPAGPQGRGDKGTGEOGDRGKIKHRGFGSLQGGPPGPPSGQPSGASGEA 977
Qy 772 GERGGPAGSP---GFQGLPGPAGPGEAGKPGQGVPGDLGAPGSPGAG 820
Db 978 GPRGPPSAGAPKDGDLNGLSPGTPGPPGRTGDAFVGPVGPVPPGPPG 1029

RESULT 11

US-10-104-889-11
; Sequence 11, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1107 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-104-889-11

Query Match 59.6%; Score 2765; DB 15; Length 1107;
Best Local Similarity 55.4%; Pred. No. 3,2e-131;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

Qy 1 GPPGPGPTGLPGPPGRRGGPGSGRFPAGDGVAGPKGAPGAGERGSPGAPGKSGPGEA---57
Db 18 GPMGSGRGLPGPPGAPGFPQGFQGPCEGPEGASGPMGRGPPGPKNGDDGEAGKP 77
Qy 58 GRPGEAGLP---GAKGLTSGPSGPP-----DGKTGPPGAGQDGRPPGPPGAP 105
Db 78 GRPGERGPPGQAGRLPGTAGLFGMKHGRGSLDGAKGDAGPAGPKGFPSPGENGAP 137
Qy 106 GQAGVMGFPKPKGAAGBPKAGER-----GVPGPPGAVGPA-----GKDGAGAG 150
Db 138 GOMGRGLPGERGRPGAPGAGPAGAGPAGNPGADGQPGAKGANGAPGIAGAPGPGARGPS 197
Qy 151 GPPGAPGAPGRRGQGP-----GPPGPPGAVGPA-----AGSPGFGPLGPA 180
Db 198 GPRSEGPQGVGPPGPPGAPAGAGPAGNPGADGQPGAKGANGAPGIAGAPGPGARGPS 257
Qy 181 GPPGEAKPQEQVPGDLGAP-----GPSGPA-----GPPGPT 213
Db 258 GPQGFPGPPGKNSGEPGAPGSKGTAKGEGPVPVQGPAGPAGEEKRGARGEPGP 317
Qy 214 GLPFPGERGSGRFPFGADGVAGPKGAPGAGERGSPGAPGKSGPGEAGLPAGK 273
Db 318 GLPFPGERGSGRFPFGADGVAGPKGAPGAGERGSPGAPGKSGPGEAGLPAGK 377
Qy 274 GLTSGPSGPDGKTGPPGAPGQDGRPPGPPGPPGARGQAGVMGFPKGAAGEPKKAGER 333
Db 378 GLTSGPSGPDGKTGPPGAPGQDGRPPGPPGARGQAGVMGFPKGAAGEPKKAGER 437
Qy 334 GVPFPAGVAGPKKAGEAGAQGPAGPAGRGGQGPAGSPGFGQLPGPAGPGEAGKP 393
Db 438 GVPFPAGVAGPKKAGEAGAQGPAGPAGRGGQGPAGSPGFGQLPGPAGPGEAGKP 497
Qy 394 GEQVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
Db 498 GEQVPGDLGAPGSPGAGERGFPGERGVQGPAGPAGPRGANGAPGNDGAKGDAGAPGAP 557
Qy 424 GERGGPSRFPGADGVAGPKGAPGAGERGSPGAPGKSGPGE-----AGRPGAGLP 474
Db 558 GSQAGPLQMGGERGAAGLPGPKDRGADGPKGADSGFGKGVGRGLTGTPTGPPAGAP 617
Qy 475 GAK-----GLTSGPSGPDGKTGPPGAPGQDGRPPGPPGARGQ-----515
Db 618 GDRGESPSGAPGPTGARGAPDRGPPGPPGAPGAGPAGGQDQPKAGEPCDAGAKGDA 677
Qy 516 -----AGVMGFPKPKGAAGEPKGAGERGVPPGPA-----VGPAGKDEAGAGQPP 561
Db 678 GPPGAPGAPGPPGPIGNVAPKAGKARGSGAGPPGATGPPGAAGRVGPPGSGNAGPPGP 737
Qy 562 GPAGP---AGEGEQAGSPGFGQLPGPAGPPGAGKPGGQGVPGDLGAPGSPGAPGE 618
Db 738 GPAGKGGKPRGEGTGPAGRPGEVPPGPPGAPGKSGFADGAPAGPTGPPQGIAGQR 797
Qy 619 GPTGLPGRGGRGSRFPFGADGVAGPKGAPGAGSGSPGAPG---KSGPGEAGRPGEA 675
Db 798 GVVGLPQGRGERGFGFLRPSGEPCKQKQSGASGERGPPGMPGLAGPPGSGREGAP 857
Qy 676 GLPGAKGLTSGPSGPDGKTGPPGAGQDGRPPGAGQAGVMGFPKGAAGEP 735


```

Db 419 GPGGPGGPGKSGEFCAGSGKDTGAKGPGVQGGPPGAGEGKRGAGEPPT 478
Qy 214 GLPGPPGRRGGGRRGPPGADVAGPKGPAGERSGPGAGPKGSFGEAGRPGCEAGLPGAK 273
Db 479 GLPGPPGRRGGGRRGPPGADVAGPKGPAGERSGPGAGPKGSFGEAGRPGCEAGLPGAK 538
Qy 274 GLTSGPSGPDGKTGTPGPPAGQDGRPPGPPGAGQAGVGMPPGPKGAAGEPGKAGER 333
Db 539 GLTSGPSGPDGKTGTPGPPAGQDGRPPGPPGAGQAGVGMPPGPKGAAGEPGKAGER 598
Qy 334 GVPGPAGVAGPAGKDEAGAGQPPGPPAGPAGERSGPGAGSGFGLPGPAGPGEAGKP 393
Db 599 GVPGPAGVAGPAGKDEAGAGQPPGPPAGPAGERSGPGAGSGFGLPGPAGPGEAGKP 658
Qy 394 GEQGVGDLGAPGSGPAGE-----PGP-----TGLPGPP 423
Db 659 GEQGVGDLGAPGSGPARGRGFPGERGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 718
Qy 424 GERGGPGRGPPGADVAGPKGPAGERSGPGAGPKGSFGE-----AGRPGCEAGLP 474
Db 719 GSQAGPLQGMPPGERGAGLPGPKDGRDAGPKGADGSGPKDGVRLGTGPIGPPGAPAG 778
Qy 475 GAK-----GLTSGPSGPDGKTGTPGPPAGQDGRPPGPPGAGQ-----515
Db 779 GDKESGSPGAPGTGARGAGPDRGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 838
Qy 516 -----AGVMPGPPGPKGAAGEPGKAGERVPPGPA-----VGPAGKDEAGAGQPP 561
Db 839 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 898
Qy 562 GPAGP-----AGRPGCEAGLPGAGPKGPAGERSGPGAGPKGSFGEAGRPGCEAGLPG 618
Db 1019 GABGSPGRDGSFAGKDRGTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1078
Qy 619 GPTGLPGPPGRRGGGRRGPPGADVAGPKGPAGERSGPGAGP-----KGSFGEAGRPGCE 735
Db 1079 GABGSPGRDGSFAGKDRGTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1138
Qy 772 GERGEGPAGSP---GFOGLPGPAGPGEAGKPGEQVPGDLGAPGPPGPPAG 820
Db 1139 GPRGPPGSAGAPGKDLNGLNGLPPIGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1190

```

```

RESULT 15
US-10-357-851-1
; Sequence 1, Application US/10357851
; Publication No. US20040151731A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; TITLE OF INVENTION: Method and Compositions Involving
; TITLE OF INVENTION: T-Lymphocyte Reactivity with Collagen in Aortic Tissue of
; TITLE OF INVENTION: Abdominal Aortic Aneurysm Patients
; FILE REFERENCE: 13376US
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-357-851-1

```

Query Match 59.6%; Score 2765; DB 16; Length 1464;

```

Best Local Similarity 55.4%; Pred. No. 3.9e-131;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;

Qy 1 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 57
Db 179 GPMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 238
Qy 58 GRPCEAGLP---GAKGLTSGPSGPP-----DKTGGPPGAGQDGRPPGPPGPPCAR 105
Db 239 GRPGERPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 298
Qy 106 GOAGVMPGPPGPKGAAGEPGKAGER-----GVPPPGAVGPA-----GKDGCEAGQ 150
Db 299 GQMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 358
Qy 151 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 180
Db 359 GPRSGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 418
Qy 181 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 213
Db 419 GPOGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 478
Qy 214 GLPGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 273
Db 479 GLPGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 538
Qy 274 GLTSGPSGPDGKTGTPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 333
Db 539 GLTSGPSGPDGKTGTPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 598
Qy 334 GVPGPAGVAGPAGKDEAGAGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 393
Db 599 GVPGPAGVAGPAGKDEAGAGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 658
Qy 394 GEQGVGDLGAPGSGPAGE-----PGP-----TGLPGPP 423
Db 659 GEQGVGDLGAPGSGPARGRGFPGERGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 718
Qy 424 GERGGPGRGPPGADVAGPKGPAGERSGPGAGPKGSFGE-----AGRPGCEAGLP 474
Db 719 GSQAGPLQGMPPGERGAGLPGPKDGRDAGPKGADGSGPKDGVRLGTGPIGPPGAPAG 778
Qy 475 GAK-----GLTSGPSGPDGKTGTPGPPAGQDGRPPGPPGAGQ-----515
Db 779 GDKESGSPGAPGTGARGAGPDRGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 838
Qy 516 -----AGVMPGPPGPKGAAGEPGKAGERVPPGPA-----VGPAGKDEAGAGQPP 561
Db 839 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 898
Qy 562 GPAGP-----AGRPGCEAGLPGAGPKGPAGERSGPGAGPKGSFGEAGRPGCEAGLPG 618
Db 899 GPAGKEGKGRGTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 958
Qy 619 GPTGLPGPPGRRGGGRRGPPGADVAGPKGPAGERSGPGAGP-----KGSFGEAGRPGCE 675
Db 959 GWVGLPGQGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1018
Qy 676 GLPAGKGLTSGPSGPDGKTGTPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 735
Db 1019 GABGSPGRDGSFAGKDRGTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1078
Qy 736 GKAGERVPPGPPGAVGPPGPKAGE-----AGAQQGPPGPP-----AGPA 771
Db 1079 GPAGARGPAGPQGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1138
Qy 772 GERGEGPAGSP---GFOGLPGPAGPGEAGKPGEQVPGDLGAPGPPGPPAG 820
Db 1139 GPRGPPGSAGAPGKDLNGLNGLPPIGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1190

```

Search completed: June 17, 2005, 15:35:47

Fri Jun 17 15:44:49 2005

us-10-658-989a-3.rapb

Page 12

Job time : 102.932 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:42:55 ; Search time 28.1036 Seconds
(without alignments)
2810.812 Million cell updates/sec

Title: US-10-658-989A-3

Perfect score: 4640

Sequence: 1 GPPGPGTGLPGRGGRG.....GEOGVGDGLGAPGSPAGG 821

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2756	59.4	1464	1 CGHUI5	collagen alpha 1(I)
2	2732.5	58.9	1453	2 S21626	collagen alpha 1(I)
3	2679	57.7	1042	1 CGCHI5	collagen alpha 1(I)
4	2477	53.4	1418	2 T45467	collagen alpha 1(I)
5	2476	53.4	1487	1 CGHU6C	collagen alpha 1(I)
6	2460	53.0	1419	2 A41182	collagen alpha 1(I)
7	2460	53.0	1487	2 B41182	collagen alpha 1(I)
8	2401	51.7	1486	1 B40333	collagen alpha 1(I)
9	2389	51.5	1492	2 A40333	collagen alpha 1(I)
10	2379	51.3	1049	1 CGBO7S	collagen alpha 1(I)
11	2377.5	51.2	1464	2 S98856	collagen alpha 1(I)
12	2370	51.1	1466	1 CGHU7L	collagen alpha 1(I)
13	2326	50.1	1496	1 CGHU2V	collagen alpha 2(I)
14	2303.5	49.6	1497	2 I49607	procollagen type V
15	2300	49.6	1373	1 A43291	collagen alpha 2(I)
16	2263.5	48.8	1366	1 CGHU2S	collagen alpha 2(I)
17	2235	48.2	671	1 CGRT1S	collagen alpha 1(I)
18	2229.5	48.0	1838	1 CGHU1V	collagen alpha 1(I)
19	2208.5	47.6	1843	2 S18803	collagen alpha 1(I)
20	2193.5	47.3	779	1 CGBO1S	collagen alpha 1(I)
21	2169	46.7	1806	1 CGBO1S	collagen alpha 1(I)
22	2161.5	46.6	886	2 I50694	collagen alpha 1(I)
23	2136.5	46.0	1414	1 S23809	collagen alpha 2(I)
24	2074.5	44.7	1027	2 S28774	collagen alpha cha
25	2071	44.6	2944	2 A54849	collagen alpha 1(I)
26	2060.5	44.4	1546	1 CGHU2E	collagen alpha 2(I)
27	2058	44.4	1691	1 S22917	collagen alpha 5(I)
28	2046	44.1	1690	1 CGHU1B	collagen alpha 4(I)
29	2036	43.9	1549	2 I48103	type VII collagen

collagen alpha 2(I)
collagen alpha 1(I)
collagen alpha 1(I)
collagen alpha 3(I)
collagen alpha 1(X)
collagen alpha 2(I)
collagen alpha 2(I)
hypothetical prote
collagen alpha 3(I)
collagen alpha 1(X)
collagen alpha 2(I)
collagen alpha 1(I)
collagen alpha 1(I)
collagen COLF1 - f
collagen alpha-4 c
collagen alpha 6(I)

30 2030 43.8 1763 2 S16366
31 1996 43.0 1669 1 CGHU4B
32 1985 42.8 1669 1 CGMS4B
33 1948.5 42.0 1670 1 CGHU3B
34 1944.5 41.9 1603 2 S23810
35 1930.5 41.6 1712 1 CGHU2B
36 1915.5 41.3 1759 2 T29351
37 1910 41.2 1758 2 T29350
38 1908.5 41.1 1752 2 A45407
39 1897 40.9 1024 2 S18251
40 1889 40.7 1707 2 A33526
41 1870.5 40.3 1775 2 A31893
42 1869 40.3 1744 2 S40991
43 1838.5 39.6 812 2 S31521
44 1833.5 39.5 1747 2 A54121
45 1832 39.5 1691 1 CGHU6B

ALIGNMENTS

collagen alpha 1(I) chain precursor - human
N:Alternate names: procollagen alpha 1(I) chain
C:Species: Homo sapiens (man)
C:Date: 12-Aug-1981 #sequence_revision 04-Oct-1996 #text change 09-Jul-2004
C:Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S11:
5269; A29439; I53466; A02852; I37247
R:D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.
Gene 67, 105-115, 1988
A:Title: Complete nucleotide sequence of the region encompassing the first twenty-five ex
A:Reference number: I60114; MUID:88329734; PMID:2843432
A:Accession: I60114
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-369, 'L', 371-589 <DAL>
A:Cross-references: UNIPROT:P02452; UNIPROT:Q14992; UNIPROT:Q16053; UNIPROT:Q13896; UNIP:
R:Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Procke
Biochem. J. 253, 919-922, 1988
A:Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human t
A:Reference number: S01143; MUID:89025644; PMID:3178743
A:Accession: S01143
A:Molecule type: mRNA
A:Residues: 1-472 <TRO>
A:Cross-references: EMBL:X07884; NID:G30015; PIDN:CAA30731.1; PID:G30016; GB:M36546; NID:
A:Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.;
Nature 310, 337-340, 1984
A:Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of
A:Reference number: A93335; MUID:84270697; PMID:6462220
A:Accession: A93335
A:Molecule type: DNA
A:Residues: 1-58, 'Q', 60-181 <CHU>
A:Cross-references: EMBL:X00820; NID:G35657; PIDN:CAA25394.1; PID:G35658
R:Rossouw, C.M.S.; Vergeer, W.P.; du Flooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.;
J. Biol. Chem. 263, 15151-15157, 1987
A:Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enha
A:Reference number: I55254; MUID:88033098; PMID:2822714
A:Accession: I55254
A:Status: translation not shown; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-45 <ROS>
A:Cross-references: GB:J02829; NID:G180387; PIDN:AAA51993.1; PID:G180388
R:Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gellinas, R.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A:Title: Regulatory elements in the first intron contribute to transcriptional control of
A:Reference number: A39943; MUID:88097389; PMID:3480516
A:Accession: A39943
A:Molecule type: DNA
A:Residues: 1-34 <BOR>
A:Cross-references: GB:J03559; NID:G180876; PIDN:AAA52052.1; PID:G553238
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

J. Biol. Chem. 260, 2315-2320, 1985
 A>Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
 A;Reference number: 155237; MUID:85130970; PMID:2857713
 A;Accession: 155237
 A>Status: translation not shown; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-34 <CH2>
 A;Cross-references: GB:M10627; NID:G180383; PIDN:AAA51992.1; PID:G553226
 R;Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
 J. Biol. Chem. 265, 6312-6317, 1990
 A>Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina
 rone, type VII.
 A;Reference number: A35233; MUID:90202908; PMID:2318855
 A;Accession: A35233
 A;Molecule type: protein
 A;Residues: 33-52 <WR>
 A;Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
 R;Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
 EMBO J. 8, 1705-1710, 1989
 A>Title: A base substitution in the exon of a collagen gene causes alternative splicing
 A;Reference number: S09400; MUID:89356643; PMID:2767050
 A;Accession: S09400
 A;Molecule type: mRNA
 A;Residues: 156-183 <WEI>
 R;Click, E.M.; Bornstein, P.
 Biochemistry 9, 4699-4706, 1970
 A>Title: Isolation and characterization of the cyanogen bromide peptides from the alpha1
 A;Reference number: A90567; MUID:71038625; PMID:5529814
 A;Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
 A;Accession: B90567
 A;Molecule type: protein
 A;Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'
 A;Experimental source: skin
 A;Note: evidence for 170-alanine
 R;Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, F
 Eur. J. Biochem. 192, 153-159, 1990
 A>Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
 A;Reference number: S11372; MUID:90382436; PMID:2169412
 A;Accession: S11372
 A;Molecule type: protein
 A;Residues: 175-187; 274-287, 'P', 289 <BAE>
 A;Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion
 R;Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
 J. Biol. Chem. 266, 21827-21832, 1991
 A>Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
 cooperative melting of intact type I collagen.
 A;Reference number: 155342; MUID:92042092; PMID:1718984
 A;Accession: 155342
 A>Status: translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 258-268; 1347-1357 <DEA>
 A;Cross-references: GB:S67495; NID:G239007; PIDN:AAB20350.1; PID:G239008
 A;Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
 R;Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
 J. Biol. Chem. 245, 5042-5048, 1970
 A>Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
 A;Reference number: A92069; MUID:71001508; PMID:4319110
 A;Accession: A92069
 A;Molecule type: protein
 A;Residues: 263-268 <MOR>
 A;Experimental source: skin
 A;Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
 R;Labhard, M.E.; Hollister, D.W.
 Matrix 10, 124-130, 1990
 A>Title: Segmental amplification of the entire helical and telopeptide regions of the ch
 A;Reference number: S15989; MUID:90326017; PMID:2374517
 A;Accession: S15989
 A;Molecule type: mRNA
 A;Residues: 281-302; 402-420; 823-843; 925-944; 1026-1045; 1143-1162 <LAB>
 R;Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Precorius, P.J.; de Vries, W.N
 Connect. Tissue Res. 29, 1-11, 1993
 A>Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
 A;Reference number: 152905; MUID:93339042; PMID:8339541

A;Accession: 152905
 A>Status: translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 342-352, 'C', 354-359 <W12>
 A;Cross-references: GB:S64717; NID:G408195; PIDN:AAB27677.1; PID:G408196
 A;Note: mutant sequence from patient with osteogenesis imperfecta
 R;Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
 Biochemistry 22, 5213-5223, 1983
 A>Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha 1
 A;Reference number: A90476; MUID:84080385; PMID:6689127
 A;Accession: A90476
 A;Molecule type: mRNA
 A;Residues: 425-1250, 'X', 1252-1328, 'S', 1330-1390, 'X', 1392-1464 <BER>
 A;Cross-references: GB:K01228; NID:G180391; PIDN:AAA51995.1; PID:G180392
 A;Note: sequence partially completed for missing nucleotides by A29439
 R;Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
 J. Biol. Chem. 260, 691-694, 1985
 A>Title: Multixon deletion in an osteogenesis imperfecta variant with increased type III
 A;Reference number: A22161; MUID:85104934; PMID:2981843
 A;Accession: A22161
 A;Molecule type: DNA
 A;Residues: 472-594, 'R', 596-607 <CH3>
 A;Cross-references: GB:K03178; GB:K03179; NID:G179612; NID:G179613; PIDN:AAA51847.1; PID
 A;Note: the authors translated the codon CGT for residue 595 as Pro
 R;Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
 Am. J. Hum. Genet. 46, 1034-1040, 1990
 A>Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained
 A;Reference number: A35336; MUID:90252792; PMID:2339700
 A;Accession: A35336
 A;Molecule type: mRNA
 A;Residues: 710-720, 'B', 722-737, 'E', 739-745 <WAI>
 A;Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
 R;Porlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes,
 Hum. Mol. Genet. 3, 2201-2206, 1994
 A>Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the co
 A;Reference number: 154365; MUID:95187161; PMID:7881420
 A;Accession: 154365
 A>Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 746-766, 'S', 768-781 <FOR>
 A;Cross-references: GB:I47667; NID:G1009093; PIDN:AAB59576.1; PID:G1009094
 R;Chessier, S.D.; Wallis, G.A.; Byers, P.H.
 J. Biol. Chem. 268, 18218-18225, 1993
 A>Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
 A;Reference number: A47426; MUID:93352646; PMID:8349697
 A;Accession: A47426
 A;Molecule type: mRNA
 A;Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CHE>
 A;Cross-references: GB:S64596; NID:G407589; PIDN:AAB27856.1; PID:G407590
 A;Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBI:136445)
 A;Note: does not represent an experimentally determined sequence but three different mut
 A;Accession: B47426
 A;Molecule type: mRNA
 A;Residues: 1179-1464 <CH4>
 A;Experimental source: normal dermal fibroblast culture
 A;Accession: C47426
 A;Molecule type: mRNA
 A;Residues: 1179-1276, 'H', 1278-1464 <CH5>
 A;Experimental source: fetal cell 86-237
 A;Accession: D47426
 A;Molecule type: mRNA
 A;Residues: 1179-1336, 1339-1464 <CH6>
 A;Experimental source: fetal cell 86-146
 A;Accession: E47426
 A;Molecule type: mRNA
 A;Residues: 1179-1387, 'R', 1389-1464 <CH7>
 A;Experimental source: fetal cell 88-251
 R;Conn, D.H.; Apone, S.; Byre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.; Nict
 J. Biol. Chem. 263, 14605-14607, 1988
 A>Title: Substitution of Cysteine for Glycine within the Carboxyl-terminal Telopeptide of
 A;Reference number: 155269; MUID:89008319; PMID:3170557
 A;Accession: 155269


```

Qy 736 GKAGERVGPAGVAGPAGKDKGE-----AGAQGGPPGP-----AGPA 771
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1079 GPAGARGPAGPQGPGRGDKGTGEQDGRGIKGHRGFSGLVQPPGSPGEGFSGASGPA 1138
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 772 GERGEQCPAGSP--GFOGLPGPAGPGCEAGKPGEQGVPGDLGAPGPSGPA 820
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1139 GPRPPGSGAGPKGDLGNLPGPIGPPGPRGTGAGPVGPPGPPGPPG 1190
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 2
S21626
Collagen alpha 1(I) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 13-Jan-1995 #sequence_revision 25-Apr-1997 #text change 09-Jul-2004
C:Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626
R:Li, S.W.; Khillan, J.; Prockop, D.J.
Matrix Biol. 14, 593-595, 1994
A:Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I
A:Reference number: S57243
A:Accession: S57243
A:Molecule type: mRNA
A:Residues: 1-1453 <LIS>
A:Cross-references: UNIPROT:P11087; EMBL:U08020; NID:G470673; PIDN:AAA88912.1; PID:G470673
R:Metzgerant, M.; Roman, D.; de Crombrugne, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A:Reference number: S16176; MUID:91274355; PMID:2054384
A:Accession: S16374
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1453 <MET>
A:Cross-references: EMBL:X57981; NID:G50484; PIDN:CAA41046.1; PID:G50485
R:French, B.T.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A:Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
A:Reference number: A23982; MUID:86137403; PMID:3841523
A:Accession: A23982
A:Molecule type: mRNA
A:Residues: 518-1128 <PRE>
A:Cross-references: GB:M14423; NID:G192261; PIDN:AAA37333.1; PID:G192262
R:Monson, J.M.; Friedman, J.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A:Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for
A:Reference number: I49559; MUID:83141374; PMID:6298597
A:Accession: I49559
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 735-1130 <RES>
A:Cross-references: GB:M17491; NID:G192263; PIDN:AAA37334.1; PID:G192264
R:Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
A:Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads to
A:Reference number: I49557; MUID:84170331; PMID:6324198
A:Accession: I49557
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-25 <RES>
A:Cross-references: GB:K01698; NID:G192246; PIDN:AAA37330.1; PID:G553881
R:Fenton, S.P.; Lamande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A:Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A:Reference number: S39789; MUID:94092741; PMID:8268229
A:Accession: S39789
A:Molecule type: DNA
A:Residues: 1-80,'E',82-105,'D',107-185,1031-1201,'G',1203-1218,'E',1220-1221,'T',1223-1-
R:Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.
Mol. Cell. Biol. 14, 5950-5960, 1994
A:Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indirect
A:Reference number: I48300; MUID:94344105; PMID:8065328
A:Accession: I48300
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-80,'E',82-105,'D',107-147 <REF>

```

Query Match	58.9%;	Score	2732.5;	DB	2;	Length	1453;
Best Local Similarity	54.7%;	Pred.	No. 5.1e-129;				
Matches	547;	Conservative	33;	Mismatches	240;	Indels	180;
Gaps	16;						
QY	1	GPPECPPTCLGPPGGB-----	RGGPGSRGFP	GADGVAGPKG	PAGRGSPGPA	48	
Db	180	GPFGAFPGQFQDPPGPPGPGSGGPMGRPGPPGPKNGDDGEAGKPGRPGRGPPGQ	239				
QY	49	GPKGSGPGEACRP-----	GEAGLPCA	KGLTGSPGSP	PDGKTGPPGAGOD	93	
Db	240	GARGLPCTAGLPQMKHHRGSGLDGAKGADGAPGPKGPGSGPCENGAPCQMGRGLPGR	299				
QY	94	GRGPPGPGCARGOAGVMGPPGPKGAGE-----	PKAGERV	PPGPPGAVG	PAGKDGSA	147	
Db	300	GRGPPGTAGANDGAVGAAGPPGTGTPGPFPGAVGAKGEAGPQARGSEGPQYVR	359				
QY	148	GAQGPFGPAGPAGERGEGP-----	AGSPGFQ	GLPGPAGP	PGGEAGKPBQ	192	
Db	360	GEPPGPGPAGAGPAGNPGADGQPGAKGANGAPGIAGAPGFFCARGSPSQPPGPPGPK	419				
QY	193	GVPGDLGAP-----	GPSGPA-----	GEPTGL	PLGPPGERGP	225	
Db	420	GNSGEPGAPGNKGDTCAGKEPGATGVQPPGPGPAGEEGKRGARPEPGSLGPPGERGCP	479				
QY	226	GSRGFGADGVAGPKGPAGERSPPGAPKPGSGEACRPCEAGLPCA	KGLTGSPGSPGD	285			
Db	480	GSRGFGADGVAGPKGSGERGAPGAPGAPKPGSGFEAGRFOEAGLPCA	KGLTGSPGSPGD	539			
QY	286	GKTGPPGAGODRPPDPGPPGARGOAGVMGPPGKGAAGEPKAGERV	PPGPAVGA	345			
Db	540	GKTGPPGAGODRPPDPGPPGARGOAGVMGPPGKGTAGEPKAGERL	PLGPPGPAVGA	599			
QY	346	GKDGEAGAGPPGAPGACGERGEPGAGSPGQGLPGPAGPPGEAGKPG	EQCVGPDGAP	405			
Db	600	GKDGEAGAGPAGPAGPAGERGEGPAGSPGQGLPGPAGPFOEAGKPG	EQCVGPDGAP	659			
QY	406	GPSGPAGE-----	PGP-----	TGLP	PPGERGPGSRGP	435	
Db	660	GPSGARGERFPGERGVQPPGPPGAPGRGNNGAPGNDGAKGDTCA	PCAPSGQAGPLQGP	719			
QY	436	GADGVAGPKGPAGERSPPGAPKPGSPB-----	AGPGEAGL	PCA	KGLTGSPGSP	486	
Db	720	GERGAAGLPKPGKDRGADGPKGADGSPKDGARGLGTGPIGPPGAGA	PDKGEAGPSGP	779			
QY	487	GPD-----	GKTGPP	PDGQDRPGPPPGARGQ-----	AGVMGFP	522	
Db	780	GPTGARGAPDRGEAGPPGAFAGPPGADCGQAGAKGEPDGTGVKGDAGP	PGPAGPAGP	839			
QY	523	GPKGAAGEPKAGERGVPPCA-----	VGPAGK	DGEAGOGPPGAP	---AGER	570	
Db	840	GPIGNVAGPKPGGAAGPPGATGFPAGAKVGPSPGSGNAGPPGP	PGVKEGKGKPR	899			
QY	571	GEQGPAGSPFQGLPGPAGPPGEAGKPGEQVPGDLGAPGSPGAGE	PGPTGLPGPPGER	630			
Db	900	GEBTGAPRGFVGPPGPPGAPAGEKSGSPGADGSPAGSPGTGPGQI	AGORGVVGLPQRCER	959			
QY	631	GQPSRGFPAGDGVAGPKGAPERSPPGAPB---	KGSPGE	AGR	PGEACLPCA	KGLTGSP	687
Db	960	GFFGLPGSPGPKGCPGSGSERGFPGPMGPPGLAGPPSGRGSG	PGABGSP	1019			
QY	688	GSPGPDGKTGPPGAGQDRPGPPGPPGARGOAGVMGPPGPKGAAGE	PKAGERGVGP	746			

```
Db 383 CSPGPDGKTPPPAGQDGRFGPAGPPGARGQAGVMGFPFGKGAAGBPKGRCRGAQCPP 442
Qy 340 CAVGPAGKDGAGAGQPPGPPAGPAGERGEQGPAGSPGFGQLPGPAGPPGKAGKPGGEQVPP 399
Db 443 CAVGAAGKDGAGAGQPPGPPGTPAGERGEQGPAGAPFGQLPGPAGPPGKAGKPGGEQVPP 502
Qy 400 GDLGARGSPGAGE-----PCP-----TGLPSPGERGPP 429
Db 503 GNAGAPGAGARGRGFFGGERGVQPPGPPQPRGANGAPNDGAKGDAGAPGPNRGP 562
Qy 430 GSRFPFGADGVAGPKPAGERGSGPP-----AGPKGSPGAGRP-----GEA 471
Db 563 GLEWMPGERGAAGLPGAKGDRGDPGPKGADGAPKDGRLGLTGPFGPPGAGAPGDKGEA 622
Qy 472 GLPGAAGLUTGSPGSPGPPGKTPGPPGA-----GDQGRGPPPP 507
Db 623 GPPGPGTARGAPGDRGEPGPPGPPGAGFPAGGADGQPKAGKGETGADAGAKGDAGPPGPA 682
Qy 508 GPPGARGQAGVMGFPFGKGAAGEPKKAGRGVPPGPPGAVGAGKDGGEAGAGQPPGAPGA 567
Db 683 GPTGAPGAGVZGAPGPKGARGSGAGPPGATGFPGAAGRVGPPGSGNTGLPFGPPGAGKZ 742
Qy 568 GE---RGQGPAGSPGFGQLPGPAGPPGKAGKPGGQGVPGDLGAPGSPGAPGPPGPTGLP 624
Db 743 GSKGPRGETGACRGECEGPPGPPGPKGKSPGADGPIGAPTPGPGQIAGQRGVWGLP 802
Qy 625 GPPCRRGSGRSGFPAGDVGAPKGPAGERGSPGAPG---KGSFGKAGRPCEAGLPGAK 681
Db 803 GQRGEPFGLPGPSGEPKQGPSKASGERGPPGPMGPPGLAGPPGKAGRGAPGAEAP 862
Qy 682 GLTSPGSGPDGKTPGPPGAGQDRPFPPPGPPGARGQAGVMGFPFGKGAAGEPKKAGRG 741
Db 863 GRDGAAGPKGDRGETGAPGAPGAPGAPGVPAGKNGDRGETGAPGAPGPPGAPG 922
Qy 742 GVPGPPGAVGAPAGKDG-----AGAQGPP-----GPAGPAGER 774
Db 923 GPAGQGRGDKGTGEGDGRMKHRCFSLGQPPGPPGAPGBOGPPSGAGPAGRPP 982
Qy 775 GEQGPAGSPGFGQLPGPAGPGEAGKPGGQGVPGDLGAPGSPGAPG 820
Db 983 GSAGAAGKDLNGLPGPIGPPGPRGRTGEVGVGPPGPPGPPG 1028

RESULT 4
T45467
collagen alpha 1(II) chain precursor [imported] - horse
N:Alternate names: type II collagen
C:Species: Equus caballus (domestic horse)
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
R:Richardson, D.W.; Dodge, G.R.
A:Submitted to the EMBL Data Library, June 1996
A:Description: Cloning of equine type II collagen and modulation of its expression in eq
A:Reference number: 222977
A:Accession: T45467
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1418 <RIC>
A:Cross-references: UNIPROT:Q28396; EMBL:U62528; PIDN:BA05773.1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
```

Query Match 53.4%; Score 2477; DB 2; Length 1418;
Best Local Similarity 50.6%; Pred. No. 2.4e-116;
Matches 500; Conservative 48; Mismatches 272; Indels 168; Gaps 17;

```
Qy 1 GPPGPGPTGLPDPGPRGSGRSGFPAGDGVAGPKPAGRGSGPPGPKGSPGKAGRP 60
Db 156 GNPSPGPGVSGMPGPPGPPGPKGPDGDRGAGPKGSGRGGPPGPPGPPGPPG 215
Qy 61 -----GEAGLPGKAGLGTSP---GSPGP-----DGKTPGPPGA--- 90
Db 216 GVKHGRGVPLDGAKEAGAPGVKSGSGSPGSGPPGPPGPPGPPGPPGPPG 275
```

```
Qy 91 QODRRPDPGPPGARGQAGVMGFPFGKGAAGEPKKAGRG-----GVPPG- 134
Db 276 GNDQPPGAPGPPGVPAGGPPGPPGAGKAGEAGTARGPEAGQPRGEPGTGSGPGA 335
Qy 135 -----PGAVGPAGKDGGEAGAGQPPGPPGAPGAGERGEQGP-----AGSRFGQGLP 177
Db 336 GNAGNPGTGTGIPGAKGSAGAPGIAGAPGFPGRGPPGPPGQATGFLGPKGQGTGFIAGFK 395
Qy 178 GPAGPPGKAGKPGGQGVPGDLGAPGSPGAPGPPGTLGPPGPPGPPGPPGPPGPPG 237
Db 396 GEQPKGEPGPPGAPGQAGPAGBEGKRGARGEPGAGVPGPPGPPGPPGPPGPPG 455
Qy 238 GPKPAGERSGPPGAPGKSPGSEAGRPGEAGLPGAKGLTGSPPGPPGPPGPPGPPG 297
Db 456 GPKGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 515
Qy 298 GRPFPFPFPARGQAGVMGFPFGKGAAGEPKKAGRGVPPGPPGAVGAGKDGGEAGAGQPP 357
Db 516 GRPFPFPFPARGQAGVMGFPFGKGAAGEPKKAGRGVPPGPPGAVGAGKDGGEAGAGQPP 575
Qy 358 GPAGPAGERGEQGPAGSPGFGQLPGPAGPPGKAGKPGGQGVPGDLGAPGSPGPPAGE---P 414
Db 576 GPAGPAGERGEQGPAGSPGFGQLPGPAGPPGKAGKPGGQGVPGPPGAGAPGLVPPRGGRFP 635
Qy 415 GPTGLP-----CPPGERGPPGSRGPPGADGVAGPKGA 447
Db 636 GERSGPAGLQAGARGLPGTGTDPKAGSAGPAGPPGAGQPPGPPGPPGPPGPPGPPG 695
Qy 448 GERSGPAGPKGSPG-----AGRPGEAGLPGKAGLGTGSPGSPGPPD 489
Db 696 GDRDGVKEGEPGAPGKDGGRGLTGPFGPPGPPGPPGPPGPPGPPGPPGPPGPPG 755
Qy 490 KGTGPPGPPGADGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 534
Db 756 GETGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 815
Qy 535 GERVPFPPPG-----VGPAGKDGGEAGAGQPPGPA---GPAGERGEQGPAGSPG 582
Db 816 GARGAQPPGATGFPGAAGRVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 875
Qy 583 GLPFPAGPPGKAGKPGGQGVPGDLGAPGSPGAPGPPGPPGPPGPPGPPGPPGPPGPPG 642
Db 876 GLQFPAGPPGKAGKPGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 935
Qy 643 GVAGPKPAGERSGPPGAPGKSPGSEAGRPGEAGLPGA-----KGLT 684
Db 936 GKQAGPAGSGDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 995
Qy 685 GSPGSPGPPGKTPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 740
Db 996 GAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1055
Qy 741 -----RGVPFPGPPGAVGPPGKDGGEAGAGQPPGPPGPPGPPGPPGPPGPPGPPGPPG 792
Db 1056 GERGLKHRGFTGLQGLPFPFPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1115
Qy 793 GPPGKAGKPGGQGVPGDLGAPGSPGPPG 820
Db 1116 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1143

RESULT 5
CCHUEC
collagen alpha 1(II) chain precursor [validated] - human
N:Alternate names: procollagen alpha 1(II) chain
N:Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collagen 2
C:Species: Homo sapiens (man)
C:Date: 28-May-1986 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C:Accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S631
7250; I37251; I37252; I37253; I37254; I55338; I59535; I61910
R:Ryan, M.C.; Sieraski, M.; Sandell, L.J.
Genomics 8, 41-48, 1990
A:Title: The human type II procollagen gene: identification of an additional protein-codi
```

R;Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage, Eur. J. Biochem. 234, 125-131, 1995

A;Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil cartilage

A;Reference number: S63514; MUID:96096730; PMID:8529631

A;Accession: S63514

A;Molecule type: protein

A;Residues: 243-261;575-590;756-763,'X',765-779 <FRA>

R;Filler, G.E.; Weis, M.A.; Polunbo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, Am. J. Hum. Genet. 56, 388-395, 1995

A;Title: An RNA-splicing mutation (G+5IVS20) in the type II collagen gene (COL2A1) in a family with osteogenesis imperfecta

A;Reference number: I38867; MUID:95150028; PMID:7847372

A;Accession: I38867

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 440,'G',442-456,'E',458-480,'P',482-509 <TILL>

A;Cross-references: EMBL:U15195; NID:9557053; PIDN:AA860370.1; PID:9557054

R;Ramirez, F.

A;Title: A mutation in the type II collagen gene (COL2A1) in a family with osteogenesis imperfecta

A;Reference number: S04892

A;Reference number: S04892

A;Accession: S04892

A;Molecule type: mRNA

A;Residues: 501-676,'A',678-783,'A',785-831,'PA',834,'F',836-1214 <RAM>

A;Cross-references: EMBL:X13783; NID:g30037; PIDN:CAA32030.1; PID:g930050

R;Viikula, M.; Peltonen, L.

A;Title: Structural analyses of the polymorphic area in type II collagen gene.

A;Reference number: S05000; MUID:89325561; PMID:2753125

A;Accession: S05000

A;Molecule type: DNA

A;Residues: 630-640,'A',642-785 <VIK2>

A;Cross-references: EMBL:X16158; NID:g29951; PIDN:CAA34278.1; PID:g1335018; PIDN:CAA34275

A;Reference number: S04309; MUID:93054548; PMID:1429602

A;Accession: S04309

A;Status: nucleic acid sequence not shown; not compared with conceptual translation

A;Molecule type: DNA; mRNA

A;Residues: 752-831,'PA',834,'F',836-1005,'K',1007-1036,'Q',1038-1052,'E',1054-1068,'T',...>

A;Cross-references: GB:L00977; NID:g180812; PIDN:AA823914.1; PID:g258774

A;Note: sequence extracted from NCBI backbone (NCBI:117273); parts of this sequence were not annotated and this publication is not cited in GenBank

A;Note: mutant sequence associated with perinatal lethal hypochondrogenesis

R;Filler, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.

A;Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual with osteogenesis imperfecta

A;Reference number: S16502; MUID:90251662; PMID:2339128

A;Accession: S16502

A;Molecule type: DNA

A;Residues: 1164-1184,'GPGKDGANGIPGPI',1185-1199 <TILL2>

A;Cross-references: EMBL:M37126; NID:g180808; PIDN:AAA52037.1; PID:g180809

A;Note: mutant sequence from a patient with spondyloepiphyseal dysplasia

R;Chen, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosfeld, F.G.; Solomon, E.

A;Title: Identification and characterization of the human type II collagen gene (COL2A1).

A;Reference number: A02858; MUID:85190534; PMID:3857598

A;Accession: A02858

A;Molecule type: DNA

A;Residues: 1032-1056,'N',1058-1068,'T',1070-1487 <CHE>

A;Cross-references: GB:J00116; NID:g180395; PIDN:AAA51997.1; PID:g180396

R;Elima, K.; Vuorio, T.; Vuorio, E.

A;Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) collagen gene

A;Reference number: A27280; MUID:88067771; PMID:2825137

A;Accession: A27280

A;Molecule type: DNA; mRNA

A;Residues: 1175-1487 <ELI>

A;Cross-references: EMBL:X06268; NID:g30096; PIDN:CAA29604.1; PID:g30097

A;Experimental source: fetal epiphyseal cartilage

R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.

A;Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.

A;Reference number: A57033; MUID:87099927; PMID:3800925
A;Accession: A57033
A;Molecule type: protein
A;Residues: 'XE',1244-1246,'N',1248,'X',1250-1265;1295-1305;1395-1408 <VAN>
A;Note: chondrocalcin identified as released collagen I(II) chain carboxyl-terminal propeptide
R;Strom, C.M.; Upholt, W.B.
Nucleic Acids Res. 12, 1025-1038, 1984
A;Title: Isolation and characterization of genomic clones corresponding to the human type I collagen alpha 1(I) chain
A;Reference number: A21733; MUID:84118798; PMID:6320112
A;Accession: A21733
A;Molecule type: DNA
A;Residues: 1245-1295 <STR1>
A;Cross-references: EMBL:X00339; EMBL:X00298; NID:g394699; PIDN:CRAA25092.1; PID:g4378975
A;Accession: B21733
A;Molecule type: DNA
A;Residues: 894-909,'PE', <STR2>
A;Cross-references: GB:K01785; NID:g30035; PIDN:CRAA25082.1; PID:g1335032
R;Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A;Title: Isolation and partial characterization of genomic clones coding for a human procollagen alpha 1(I) chain
A;Reference number: A24561; MUID:861041139; PMID:3002437
A;Accession: A24561
A;Molecule type: DNA
A;Residues: 1296-1358 <NUN2>
A;Cross-references: GB:M12048; NID:g180017
A;Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
A;Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with the
R;Sangiorzi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez,
Nucleic Acids Res. 13, 2207-2225, 1985
A;Title: Isolation and partial characterization of the entire human pro alpha 1(II) collagen
A;Reference number: I37249; MUID:85215609; PMID:2987845
A;Accession: S59491
A;Molecule type: DNA
A;Residues: 7-28,'R', 99-114;541-578;786-802;1055-1056,'N',1058-1068,'T',1070-1109;1200-1
A;Accession: I84453
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 7-28 <SAN2>
A;Cross-references: GB:M23759; NID:g180845; EMBL:X03320; GB:M24938; NID:g30104
A;Note: the GenBank PID is based on an incorrect reading frame
A;Accession: I37250
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 541-560 <SAN3>
A;Cross-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CRAA26227.1; PID:g929621
A;Accession: I37251

Query Match	53.4%	Score	2476;	DB 1;	Length	1487;			
Best Local Similarity	50.9%;	Pred. No.	2.7e-116;						
Matches	503;	Conservative	45;	Mismatches	272;	Indels	168;	Gaps	17;

Qy	1	GPCEPPTGLPGPBERGGPSRGFFCAGDVAGPKGPAGERGSPGAPGKSPCEAGRP	60
Db	225	GNPEGEPGVSGPMGRPGPPGPKPGDDGEAGKPKGAGERGPPGPGQARGFPPTGLP	284
Qy	61	-----GBAGLPCAKGLTGSP--GSPGP-----DGKTGPPGPA---90	
Db	285	GVKCHRGYPGLDGAKGEAGACVKGESGPGMGPRGLPCEKRGRTGPAAGAACR	344
Qy	91	GDGRPGPPGPGARGQAGVMGFP-----GPKGAAGBFGKAGERGVPPGPAVGA	141
Db	345	GNDQPGPAGPPGVPAGGPGFFGAPGAKGEAGTARGPEGAQGPGRGPTGSPGPA	404
Qy	142	-----GKDGEAGAGPPGAPGAGERGEQGPAGSPGFGCLPGPNAGPPGEAGKPGEGQVP	195
Db	405	GASGNPGTDGTPGAKGSAGAPCIAGAPGFPGRGPPDPQAGATPLGPKQTKCPGIAGFK	464
Qy	196	GD---LGAPGSPGABGPTGLP-----GPPREGGSPSGFFPGADGVA	237
Db	465	GEQFGKEPGPAGPQAGPFGPAGEEGKRGARGEPGVGPIGPPEERKAGPNTGFFPQDGLA	524
Qy	238	GPKGPAGERGSPGAPGKSGPCEAGRLPFAKGLTGTSGPSGPPGPKGTGPPGPAGOD	297

Db QY Db QY Db QY Db QY Db QY Db QY Db QY Db QY Db QY Db QY Db QY Db QY

525 GPKGAPGERGPSGLGKPGKANGDPRGPPLPGAKGLTGRPDAGPQKGVGSPSGAPGED 584
298 GRPPPPGGPARGAQGMVGFPPFKGAAGSPGKAGERGVPFGPAVGPAKGDSGAGAQQPPP 357
585 GRPPPPGGPARGAQGMVGFPPFKGANGEPGKAGEKGLPCAPLGRLLPGKDGSTGAEGPP 644
358 GTAGPAGERGEOPAGSPFQGLPGPAGPBGAEKDEOGVPCDLGAPGSPGAGE----- 413
645 GPAGPAGERGEOPAGSPFQGLPGPFPFGGKGKGDGVGPEAGAPGLVGRGRGPF 704
414 -----PQPTGLP-----GPPGERGGPSRSGFPFGADGVAGVKGPA 447
705 GERGSFGAOLQGPRCLPTGTGDPKGASGPAGPGAQPPQLQGWPERGAAGIATGPK 764
448 GERGS-----PGPAGPKGSPGEACRPGEAGLPKAAGLTGSP 483
765 GDRGDVGEKKPGCAPKCGKGRGLTGIPGPPAGANGEKEGVPPGPAGSAGARGAPGER 824
484 GSFGPDGKTTPGPAGQDGRPPGPPGARGOAGVMGPPGPKAAAGEP-----GKA 534
825 GETGPPGTSGIAGPPGADQPCKAQEQEAGQKDAGACPQGPSGAPFPQGTGVTGPK 884
535 GERGVGPPPCA-----VGPAGKDXBAGAQGPPGPA---GPAGERGEQGPAGSPGFQ 582
885 GARGAQPPGATCFPCAAGRVPCCNGNPPGPPGPGSKDKGPKARGDSGPPGRAGEP 944
583 GLPPAPPPGERAGKPGEQGVGDLGAPGSPGAPAGEPPTGLPPOPBERGPGGSRGPPGAD 642
945 GLQCPAGPPCEKEPEDDDPSGAEGSPGQGLAGRGIVGLPQQRGERGPPGLPGPSGEP 1004
643 GVAGPKGPAGERSGCPAGPKSGPGEAGRPGEAGLPCA-----KGLT 684
1005 GQQAAPGAGDGRGPPGVFPGLTGPAGBPBRGSCADGPPGRDGAAGVKGRGETGAV 1064
685 GSPGSPGDPKGTGPPGF-----AGODGRFGPPGPCARGAQGMVGPFGKGAAGEP 735
1065 GAPGAPFPSPGAPGAPTCKQDGRGEGAAGQPMWPSGAPAGARCIQGPQGRGDKGEAGEP 1124
736 GK---AGERGVPPGCAVGPAGKDGAGAACQPPGPAAGERGEREQGPAGSPFGQLPGPA 792
1125 GERGLKHGRFTGLQGLPFPSPGSDQASGPAGSPGRGPPGVPSGKDGANGIPGPI 1184
793 GPPGEAGKPBEOQVPCDLGAPSPSPAG 820
1185 GPPGPRGRSGETGAPPGNPFGPPG 1212

RESULT 6
A41182
collagen alpha 1(II) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
R;Accession: A41182; A44885
R;Metaeairanta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure,
A;Reference number: A41182; PMID:91358489; PMID:1885613
A;Accession: A41182
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-1419 <MET>
A;Cross-references: GB:M65161
R;Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
Development 111, 945-953, 1991
A;Title: Expression of the mouse alpha 1(I) collagen gene is not restricted to
A;Reference number: A44885; PMID:91347939; PMID:1879363

RESULT 6

A41182 collagen alpha 1(II) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999

C:Accession: A41182; A44885

R:Metaauthor, N.; Toman, D.; de Crombrughe, B.; Vuorio, E.

J. Biol. Chem. 266, 16862-16869, 1991

A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure and expression of complementary DNA.

A:Reference number: A41182; MUID:91358489; PMID:1685613

A:Accession: A41182

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-1419 <MET>

A:Cross-references: GB:M65161

R:Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.

Development 111, 945-953, 1991

A:Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to the developing embryo.

A:Reference number: A44885; MUID:91347939; PMID:1879363

A:Accession: A44885

A:Molecule type: DNA

A:Residues: 1-28 <CHE>

A:Cross-references: GB:S63190; NID:G234368; PIDN:AAB19627.1; PID:G234369

A:Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBIPI:63192)

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal

C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein, trimer
F;1191-1419/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

```
Query Match          53.0%; Score 2460; DB 2; Length 1419;
Best Local Similarity 50.6%; Pred. No. 1.6e-115;
Matches 500; Conservative 49; Mismatches 271; Indels 168; Gaps 17;

QY 1 GPPGPGPTGLPGPPGPGSGSRFPFGADGVAGPKGAGERSGPPAGPKGSPGEAGRP 60
Db 157 GNPGEFPGVSGPMGPRPPGPKGDDGAGKPKSGERGLPGFGAGRGFTGTGLP 216
QY 61 -----GEAGLPGAKGLTGSP-----GSPGP-----DGKTPPGPA--- 90
Db 217 GYVGHRCYPCLDGAKGAGAVGVKSGSPGNGSGPGRGLPGERGTGPAGAAGAR 276
QY 91 GQDGRPPGPPGARGQAGVGMGFP-----GPKGAAGPPGKAGRGVPPGPGVAGPA 141
Db 277 GNDGQPGGAPGPPGVPAGGPGFPCAGKAGGFTGARGFEGAGQGRGEPNGPSGPA 336
QY 142 GKDGE-----AGAQPPGPPGAGRGEGQP-----AGSPFGOGLP 177
Db 337 GASGNPPTDIPGAKGSAGAPGAGAPFGPRGPPGPGQATGTLGPKGQAGEPPIAGFK 396
QY 178 GPAGPFGAGKPGEGQVQDLGAPGSPGAGEPPTGLPGPPGRRGPGSGRPGADGVA 237
Db 397 GDQGPKEGTGAPGAGPAGGPKGARGEPGAGPIGPPGRRGAPGNRGPFGDGLA 456
QY 238 GPKGAPGRRGPPGAPGPKSGPGEAGRPGEAGLPGKAGLTSPPSGDGTGTPPGAGD 297
Db 457 GPKGAPGRRGPPGAPGPKSGPGEAGRPGEAGLPGKAGLTSPPSGDGTGTPPGAGD 516
QY 298 GRPFPFPAGARGQAGVGMGFPFGPKGAGGPKGAGRGVPPGPGVAGVAGKAGAGQPP 357
Db 517 GRPFPFPAGARGQAGVGMGFPFGPKGAGGPKGAGRGVPPGPGVAGVAGKAGAGQPP 576
QY 358 GPAGPAGRRGPPGAPGSPFGQGLPAGPAGGPKGAGRGVPPGPGVAGVAGKAGAGQPP 413
Db 577 GPGSPAGRRGPPGAPGSPFGQGLPAGPAGGPKGAGRGVPPGPGVAGVAGKAGAGQPP 636
QY 414 -----PGTGLP-----GPPGRRGPPGSRGPPGADGVAGPKGPA 447
Db 637 GERSGPAGLQGRGLPGTGTGDPKGAAGPDPGPAQGPGLQMPGERGAAGIAGPK 696
QY 448 GERSGPAGPKGSPGE-----AGRPGEAGLPGKAGLTSPPSGDGTGTPPGAGD 489
Db 697 GDRDVGKEGPEAGPKDGGRLTGTPTGPPGAGANGKEGVEGPPGSGSTGARGAPGEP 756
QY 490 GKTGPPGAGQDGRGPPGPPGAR-----GQAGVGMGFPKGAAGEP-----GKA 534
Db 757 GETGPPGPAFAGPPGADGQPGAKGQGEAGQKGDAGAPGQPGSPGAPGPGTGTGPK 816
QY 535 GERGVPPGPA-----VGPAGKDGAGAGQPPGPA---GPAGERGQPGSPGQF 582
Db 817 GARGAQPPGATGPPGAGRGVPPGANGNPPGAPGPKAGKDGKGVGSDGPPGRAGDP 876
QY 583 GLPGPAGPGEAGKQGEQVQDLGAPGSPGAGEPPTGLPGPPGRRGPGSGRFPAGD 642
Db 877 GLOGPAGAPGKEGEPDGDPSGLDGPQGLAGQGIIVGLPQGRGERGFFGLPFPSPSE 936
QY 643 GVAGPKGAPGERSGPPGPKSGPGEAGRPGEAGLPCA-----KGLT 684
Db 937 GKQAGPAGSDRGPPGVPGLTGPAGEPREGSPGADGPPGRDGAAGVKGDRGETCAL 996
QY 685 GSPGSPGPGKTPGPPG-----AGQDGRGPPGPPGARGQAGVGMGFPKGAAGEP 735
Db 997 GAFGAPGPPGSPGAPGTPKQDGRGAGAGQPMGSPGAGARGTAGPGPGKDGESGEQ 1056
QY 736 GK---AGERGVPPGAVGAPKAGDEAGAGQPPGAPGAGRGQGPAGSPFGOGLPCPA 792
Db 1057 GERGLXHRGFTGLQGLPFPSPGSDQAGSAGPSPGRRPFPVGPSPGKDSNGIIPGI 1116
QY 793 GPPGAGKPGEGQVQDLGAPGSPGAG 820
```

Db 1117 GPPGPRGRSGETGVPVPGSPGPPGPPG 1144

RESULT 7
B41182

collagen alpha 1(III) chain precursor (long splice form) - mouse
C;Species: Mus musculus (house mouse)

C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 09-Jul-2004
C;Accession: B41182

R;Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

J. Biol. Chem. 266, 16862-16869, 1991

A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and

A;Reference number: A41182; MUID:91358489; PMID:1885613

A;Accession: B41182

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-1487 <MET>

C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer;

F;33-91/Domain: von Willebrand factor type C repeat homology <VMC>

F;1259-1487/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match

Best Local Similarity 50.6%; Score 2460; DB 2; Length 1487;

Matches 500; Conservative 49; Mismatches 271; Indels 168; Gaps 17;

QY 1 GPPGPGPTGLPGPPGPGSGSRFPFGADGVAGPKGAGERSGPPAGPKGSPGEAGRP 60

Db 225 GNPGEFPGVSGPMGPRPPGPKGDDGAGKPKSGERGLPGFGAGRGFTGTGLP 284

QY 61 -----GEAGLPGAKGLTGSP-----GSPGP-----DGKTPPGPA--- 90

Db 285 GVKGHRGYPGLDGAKGAGAPGVKSGSPGNGSGPGRGLPGERGTGPAGAAGAR 344

QY 91 GQDGRPPGPPGARGQAGVGMGFP-----GPKGAAGPPGKAGRGVPPGPGVAGPA 141

Db 345 GNDGQPGGAPGPPGVPAGGPGFPGAPGAKGAGTGTARGFEGAGQGRGEPNGSPGPA 404

QY 142 GKDGE-----AGAQPPGPPGAGRGEGQP-----AGSPFGOGLP 177

Db 405 GASGNPPTDIPGAKGSAGAPGAGAPFGPRGPPGPGQATGTLGPKGQAGEPPIAGFK 464

QY 178 GPAGPFGAGKPGEGQVQDLGAPGSPGAGEPPTGLPGPPGRRGPGSGRFPAGD 489

Db 465 GDQGPKEGTGAPGAGPAGGPKGARGEPGAGPIGPPGRRGAPGNRGPFGDGLA 524

QY 238 GPKGAPGRRGPPGAPGPKSGPGEAGRPGEAGLPGKAGLTSPPSGDGTGTPPGAGD 297

Db 525 GPKGAPGRRGPPGAPGPKSGPGEAGRPGEAGLPGKAGLTSPPSGDGTGTPPGAGD 584

QY 298 GRPFPFPAGARGQAGVGMGFPFGPKGAGGPKGAGRGVPPGPGVAGVAGKAGAGQPP 357

Db 585 GRPFPFPAGARGQAGVGMGFPFGPKGAGGPKGAGRGVPPGPGVAGVAGKAGAGQPP 644

QY 358 GPAGPAGRRGPPGAPGSPFGQGLPAGPAGGPKGAGRGVPPGPGVAGVAGKAGAGQPP 413

Db 645 GPGSPAGRRGPPGAPGSPFGQGLPAGPAGGPKGAGRGVPPGPGVAGVAGKAGAGQPP 704

QY 414 -----PGTGLP-----GPPGRRGPPGSRGPPGADGVAGPKGPA 447

Db 705 GERSGPAGLQGRGLPGTGTGDPKGAAGPDPGPAQGPGLQMPGERGAAGIAGPK 764

QY 448 GERSGPAGPKGSPGE-----AGRPGEAGLPGKAGLTSPPSGDGTGTPPGAGD 489

Db 765 GDRDVGKEGPEAGPKDGGRLTGTPTGPPGAGANGKEGVEGPPGSGSTGARGAPGEP 824

QY 490 GKTGPPGAGQDGRGPPGPPGAR-----GQAGVGMGFPKGAAGEP-----GKA 534

Db 825 GETGPPGPAFAGPPGADGQPGAKGQGEAGQKGDAGAPGQPGSPGAPGPGTGTGPK 884

QY 535 GERGVPPGPA-----VGPAGKDGAGAGQPPGPA---GPAGERGQPGSPGQF 582

Db 885 GARGAQPPGATGTFGAAGRVGPPGANGNPGPAGPPGAGKDGKPKGVGRDGGPPGACDP 944
 Qy 583 GLPDPAGPBGAGKCEGQVPGDILGAPDPSGAPGPGTGLPGPPGEGGSGRSGFPDAD 642
 Db 945 GLQGPAGAPGKSGEPGDPGLDGGPSLGGPQGLAGQGVGLPGQGRGRTFGLPGPSGEP 1004
 Qy 643 GVAGPKGPAGERSGCPAGPKSGPCEAGRPCEAGLPGA-----KGLT 684
 Db 1005 GKQAPGASGDRGPPGVPVPGTLTPAGEPGREGSFGADGPPGRDGAAGVKGDRGTAL 1064
 Qy 685 GSPGSGPDGKTGPPGP-----AGQDRGPPGPPGARGQAGVGMFFGPKGAAGER 735
 Db 1065 GAPGAPGPPGSPGAPGTPGKQDRCEAGAQAQPMGSPGAGARGIAGPQGRGDKGESGEQ 1124
 Qy 736 GK---AGRGVPPGCAVGPAGKDEAGCAQAQPPGAPGAGERGEGGAGSGPQGLPCPA 792
 Db 1125 GERGLKGRGFTGLQGLPGPQPSDQOGASGAPGSPGPPGVPVGPSPGDKGSGNIPGI 1184
 Qy 793 GPPGAGKPGEGQVPGDILGAPGSPGAG 820
 Db 1185 GPPGPRGRSGTGTGVPGSGPSPGPPG 1212

RESULT 8
 B40333
 collagen alpha 1(II) chain precursor - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
 C:Accession: B40333
 R:Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
 J. Cell Biol. 115, 565-575, 1991
 A:Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
 A:Reference number: A40333; MUID: 92011898; PMID: 1918153
 A:Accession: B40333
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1486 <SUA>
 A:Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; GB:M63595
 C:Superfamily: collagen alpha 1(i) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
 F:37-96/Domain: von Willebrand factor type C repeat homology <VWC>
 F:1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 51.7%; Score 2401; DB 1; Length 1486;
 Best Local Similarity 48.9%; Pred. No. 1.4e-112;
 Matches 493; Conservative 48; Mismatches 276; Indels 192; Gaps 16;

Qy 1 GPPGPRGPTGLPGPPGEGGSGRSGFPAGDGVAGPKGP-----AGER 42
 Db 209 GPRGPPGTGAPPGQFQGNPGEPGEPGAGGPMGRGPPGSGKPGDDGKAGKPKSGER 268
 Qy 43 GSPGAPGPKSGPGEAGRP-----GEAGLPKAK---GLTSGSPGPP--- 80
 Db 269 GPPGQAGRGFTGLPGVKGHRGYPGLDGAKEAGAAKAGEGATGEAGSPGPMGR 328
 Qy 81 -----DKTGPAGQDGRGPPGPPGARGQAGVGMFFGPK 117
 Db 329 GLPGRGRGPGSSAAGARGNDGLPGPAGPPGVPAGAPGPPGADPKSGEAGTGAARGE 388
 Qy 118 GAAGEPKAGRGVPPGCAVGPAGKDEAGCAQAQPPGAPGAGERGEGGAGSGPQGLP 177
 Db 389 GAQGRGSGTGPSPGAGASNPCTDGIKAGSGSGFGIAGAPFPFGPPGPPQAG 448
 Qy 178 GPAGPBGAGKP-----GEQVPGDILGAPGS---GPA-----GEPGPTGLP 219
 Db 449 GPLGPKGTGDPGAVGFKGQCPKEIGSAGPQAGPAGBEGKKGAGFGAAGNPP 508
 Qy 220 GRRGPGSGRPPGADGVAGPKGPAGERSGPPGAPKSGPGEAGRPCEAGLPAKGLTSP 279
 Db 509 GERGAPGNRGPQGLAGPKGAPGERGVPLGGPKGNGNDPRGEPGLPGARGLTGRP 568
 Qy 280 GSPGPDGKTGPPGAGQDGRGPPGPPGARGQAGVGMFFGPKAGRGKAGRGVPPPP 339

Db 569 GDAGPQKVGPSGASGEDRGPPGPPQARGQPCVMGFPFGPKGANGBPKAGKGLVGP 628
 Qy 340 GAVGPAGKDGEGAGAQGPAGPAGEREGEQPAGSPGQGLPGPAGPPGEGAKGEGEQVP 399
 Db 629 GLRGLPKDGTGSGQPNGPAGPAGEREGEQPAGSPGQGLPGPAGPPGEGAKGEGEQVP 688
 Qy 400 GDLCAVPSPGAGS---PGPTGLPGPGERGCGSRGPPGADGVAGPKGPAGERSGCPA 456
 Db 689 GEAGAPGLVGRGRGPPGEGSSGQQLQPRGLPTGPTGTDGPKGASGSPGNGAQGP 748
 Qy 457 GPKSGPGEAGRPBAGLPGAKGLTSGSPGPDGKTGPPGAGQDGRGPPGPPGPPGCAQQA 516
 Db 749 GLQMPGERGAAGISGPKDGRDGTGKGPESKDSKSGRLTGTGTPGPPAGPANGKES 808
 Qy 517 GVMGPPGPKGAAGPFGKAGRGVPPGAGVGPAGKD-----GEAGAQGP 561
 Db 809 GPSGPPGIVGARGAPGRGNGPPGPFAGFAGPPGADGQSLKDGSGSQKGDAGAPGP 868
 Qy 562 GPA-----GPAGERGEGPAGSPGQGL-----PGPAGPGEAGKP 597
 Db 869 GPSGAPGPGTGVFGPKGARGAQAGATGTPGAAGRVGTPGNGNPPGPPGPPGSAK 928
 Qy 598 GEQGVPGDILGAP-----GPSGAPGPPGPTGL----- 623
 Db 929 GPKGVGDAGPPGRAGDGLQGAAGAPCEKGEPCEDGSPGDPGPPGQGLSGQRIVGLP 988
 Qy 624 -----PGPPGRRGPPGSRGFPAGDGVAGPKGPAGERSGPPGAPKSGPGEAGRPGE- 674
 Db 989 GQGRGFPGLPGPSGPPGKQGGSSGDRGPPGPV---GPPGLTSGSGEPGEGNPGSD 1045
 Qy 675 -----AGLPKAGLTSKSPSGPDGKTGPPGAGQDGRGPPGPPG 714
 Db 1046 GPPGRDGTGIKDRGTGTLGAPGAPGAPGAPGVSPTGKQGRGSGGPPGLGSPGA 1105
 Qy 715 GARGQAGVMPGPPGKGAAGEPGKAGE---RGVPPGPPGAVGPKGDKGAGAGPAGPAGPA 771
 Db 1106 GARGLAGPQPRGDKGAGEAGRGKGRGFTGLQGLPGPPGSDGQATGAPGAPGR 1165
 Qy 772 GEREGQSPAGSPGQGLPGPAGPGEAGKPGEQVPGDILGAPGSPGAG 820
 Db 1166 GPPGVPGSPGKSGNSIGPTGPPGPRGSGTGPSPGPPGQPPGPPG 1214

RESULT 9
 A40333
 collagen alpha 1'(II) chain precursor - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
 C:Accession: A40333
 R:Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
 J. Cell Biol. 115, 565-575, 1991
 A:Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
 A:Reference number: A40333; MUID: 92011898; PMID: 1918153
 A:Accession: A40333
 A>Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-1492 <SUA>
 A:Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; GB:M63596
 A>Note: This sequence is presented as substitutions relative to another sequence in a file
 as they replace; the appropriate interpretation of the sequence figure was reconstructed
 C:Superfamily: collagen alpha 1(i) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
 F:37-96/Domain: von Willebrand factor type C repeat homology <VWC>
 F:1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 51.5%; Score 2389; DB 2; Length 1492;
 Best Local Similarity 50.6%; Pred. No. 5.7e-112;
 Matches 477; Conservative 59; Mismatches 275; Indels 132; Gaps 14;

Qy 1 GPPGPRGPTGLPGPPGEGGSGRSGFPDAD-----GVAGPKGPAGERSGPPGAPK 51
 Db 272 GPPGQAGRGFTGTPGLPGVKGHRGYPGLDGGKGAAGAKGSGASGAGAGPMPGR 331

Db 758 GP---RGPAIGPPGP---AGQPDCKEGGSGPLGIAGPRGCGRGGHGGPPGAGFP 811

QY 565 -----GPAGRGQGPAGSPGQGLPGPAGPGE---AGKPGGQGVPGDGLGAPGPGSPA 615

Db 812 GAPGNGEPGAKGERGAPGKGEQGGPPGAPGPTGSSGPAGPAGPPGQGVKGERGSGPPGTA 871

QY 616 GEPGPTGLPGPPRGGPGSGGFFGACDVAGPKPAGRGSGPGPAGPKGSPGAGRPGE- 674

Db 872 GPPGGRGLPGPPGNNGNPPGPGGAPGKDGPPGAGNSGSPGNFGIAGPKGADGAPGK 931

QY 675 -----AGLPCAAGLTCGSPGSPGPDGKTGP-----PGPAGQDGRP 708

Db 932 GPPCAQGPSPGPIGLIAGLTGARGLAGPPGMPGPRSPGPGQIGKSGHPPGASGNGER 991

QY 709 GPPGP-----PGARGQAGVMGFPKPKGAAGE---PGKAGRGVPGPP 747

Db 992 GPPPGQGLPGQGTAGTGFGRDNGSDQPGKGDGKDRBNGSGFAGPAGPGHPP 1051

QY 748 GAVGPAGKDG------AGAGPPGAPGAGRGEGQGPAGSPGFGQGLPGPA 792

Db 1052 GPVPSGSGDRGTGPAGPSGAPGAPGAPGQFGRGDKGTGTRGSGNGIKGHRGP 1111

QY 793 GPPGEAGKPGGQGVPGDGLGAPGSPGAP 820

Db 1112 GNPFGPGSGAAGHGAIGSPGAPGPG 1139

RESULT 12

CGHUTL

N;Collagen alpha 1(III) chain precursor - human

C;Species: Homo sapiens (man)

C;Date: 24-Apr-1984 #sequence revision 01-Sep-1995 #text change 09-Jul-2004

C;Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90399

R;Prockop, D.J.

submitted to the EMBL Data Library, February 1989

A;Reference number: S05272

A;Accession: S05272

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1240, 'V', 1242-1466 <PRC>

A;Cross-references: UNIPROT:P02461; EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058

R;Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.

Biochem. J. 260, 509-516, 1989

A;Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of human

ences.

A;Reference number: S04642; MUID:89350838; PMID:2764886

A;Accession: S04642

A;Molecule type: mRNA

A;Residues: 1-1196 <ALA>

A;Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058

A;Note: the complete sequence is not shown

R;Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.

Gene 78, 255-265, 1989

A;Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene

A;Reference number: PE0011; MUID:89378752; PMID:2777083

A;Accession: PE0011

A;Molecule type: DNA

A;Residues: 1-176 <BEN>

A;Cross-references: GB:M26939; NID:g180813; PIDN:AAAS2040.1; PID:g180814

R;Toman, P.D.; Ricca, G.A.; de Crombrughe, B.

Nucleic Acids Res. 16, 7201, 1988

A;Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre

A;Reference number: S01726; MUID:88303360; PMID:3405773

A;Accession: S01726

A;Molecule type: mRNA

A;Residues: 1-170 <TOM>

A;Cross-references: EMBL:X07240; NID:g30060; PIDN:CAA30229.1; PID:g30061

A;Note: the authors translated the codon CAG for residue 154 as His

R;Janeczko, R.A.; Ramirez, F.

Nucleic Acids Res. 17, 6742, 1989

A;Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.

A;Reference number: S04887; MUID:89386015; PMID:2780304

A;Accession: S04887

A;Molecule type: mRNA

A;Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634, 'A';

A;Cross-references: EMBL:X15332; NID:g29545; PIDN:CAA33387.1; PID:g930045

A;Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide

R;Seyer, J.M.; Kang, A.H.

Biochemistry 16, 1158-1164, 1977

A;Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptides

A;Reference number: A90399; MUID:77134724; PMID:557335

A;Accession: A90399

A;Molecule type: protein

A;Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>

A;Experimental source: liver

A;Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose

R;Seyer, J.M.

submitted to the Atlas, December 1977

A;Reference number: A94562

A;Accession: A94562

A;Molecule type: protein

A;Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>

A;Experimental source: liver

A;Note: author submitted corrections to A90399

R;Milwicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.

Am. J. Hum. Genet. 53, 62-70, 1993

A;Title: Parental somatic and germ-line mosaicism for a multigene deletion with unusual

fepring.

A;Reference number: I51868; MUID:93304430; PMID:8317500

A;Accession: I51868

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 186-194 <MIL>

A;Cross-references: GB:S62925; NID:g386425; PIDN:AAD13937.1; PID:g4261637

R;Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.

Biochem. J. 311, 939-943, 1995

A;Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3A

A;Reference number: S59511; MUID:96067614; PMID:7487954

A;Accession: S59511

A;Molecule type: mRNA

A;Residues: 302-423 <CHI>

A;Cross-references: GB:S79877; NID:g1195576; PIDN:AAB35615.1; PID:g1195577

R;Seyer, J.M.; Kang, A.H.

Biochemistry 17, 3404-3411, 1978

A;Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBR

A;Reference number: A90414; MUID:79000343; PMID:687591

A;Accession: A90414

A;Molecule type: protein

A;Residues: 399-675, 'N', 677-727 <SEY3>

A;Experimental source: liver

R;Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.

J. Biol. Chem. 266, 5256-5259, 1991

A;Title: G to T transversion at position +5 of a splice donor site causes skipping of the

A;Reference number: I55349; MUID:91161621; PMID:1672129

A;Accession: I55349

A;Status: translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 537-605 <LER>

A;Cross-references: GB:M59312; NID:g180815; PIDN:AAAS2041.1; PID:g180816

R;Seyer, J.M.; Mainardi, C.; Kang, A.H.

Biochemistry 19, 1583-1589, 1980

A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from tyf

A;Reference number: A90438; MUID:80198282; PMID:6246925

A;Accession: A90438

A;Molecule type: protein

A;Residues: 728-895, 'A', 897-964 <SEY4>

A;Experimental source: liver

R;Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan,

J. Biol. Chem. 265, 17070-17077, 1990

A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping and

A;Reference number: A38303; MUID:91009133; PMID:2145268

A;Accession: A38303

A;Molecule type: mRNA

A;Residues: 861-1015 <COL>

A;Cross-references: GB:J05617; GB:M55603; NID:g180878; PIDN:AAB59383.1; PID:g1

A>Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syndrome.
 A:Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>
 A:Cross-references: EMBL:X06700; NID:G30053; PID:CAA29886.1; PID:G30054
 R:Sever, J.M.; Kang, A.H.
 Biochemistry 20, 2621-2627, 1981
 A:Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from type I procollagen
 A:Reference number: A90446; MUID:81208139; PMID:7016180
 A:Accession: A90446
 A:Molecule type: protein
 A:Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-1160
 A:Experimental source: liver
 R:Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Myer, J.
 Nucleic Acids Res. 12, 9383-9394, 1984
 A:Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollagen
 A:Reference number: A93551; MUID:85087944; PMID:6096827
 A:Accession: A93551
 A:Molecule type: mRNA
 A:Residues: 1065-1155, 'P', 1157-1466 <LOI>
 A:Cross-references: EMBL:X01655; EMBL:X01742; NID:G29584; PID:CAA25821.1
 R:Miskulin, M.; Daigleish, R.; Klueve-Beckerman, B.; Rennard, S.I.; Tolstoshev, P.; Brantley, J.E.
 Biochemistry 25, 1408-1413, 1986
 A:Title: Human type III collagen gene expression is coordinately modulated with the type I collagen gene
 A:Reference number: I52393; MUID:86187804; PMID:3754462
 A:Accession: I52393
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1161-1200 <MIS>
 A:Cross-references: GB:M13146; NID:G180415; PID:AAA52003.1; PID:G180416
 R:Emmanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
 A:Title: Human alpha 1(I) and alpha 2(V) procollagen genes are located on the long arm of chromosome 12
 A:Reference number: I59025; MUID:85216505; PMID:3858826
 A:Accession: I79359
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1165-1196 <EMA>
 A:Cross-references: GB:M11134; NID:G180417; PID:AAA52004.1; PID:G180418
 R:Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
 J. Biol. Chem. 260, 4357-4363, 1985
 A:Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. B
 A:Reference number: A92516; MUID:85157600; PMID:2579949
 A:Accession: A92516
 A:Molecule type: DNA
 A:Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>
 A:Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10797; GB:M10798
 A:Experimental source: liver
 A>Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given in the text are incorrect
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit are 5-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently C-glycosylated.
 C:Genetics:
 A:Gene: GDB:COL3A1
 A:Cross-references: GDB:118729; OMIM:120180
 A:Map position: 2q31-2q31
 A:Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3; 1419/3
 A>Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Danlos syndrome
 C:Complex: type III collagen is a homotrimer of monomers initially linked by disulfide bridges. The length of the monomer is 1018 amino acids. The monomer is formed from lysine and alllysine
 C:Function:
 A:Description: structural component of extracellular fibrous polymer that maintains integrity of connective tissue
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; C:Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-153/Domain: amino-terminal propeptide #status predicted <PRO>
 F:31-91/Domain: von Willebrand factor type C repeat homology <VWC>
 F:154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>

F:154-167/Region: amino-terminal nonhelical telopeptide
 F:168-1196/Region: helical
 F:1091-1093/Region: cell attachment (R-G-D) motif
 F:1197-1221/Region: carboxyl-terminal nonhelical telopeptide
 F:1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>
 F:1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCH>
 F:24/Modified site: pyrocollagen carboxylic acid (Gln) (in mature form) #status predicted
 F:153-154/Cleavage site: Pro-Gln (procollagen N-endoropeptidase) #status predicted
 F:154/Modified site: pyrocollagen carboxylic acid (Gln) (in mature form) #status predicted
 F:161-1212/Modified site: allysine (Lys) #status predicted
 F:263-284, 860, 977, 1106/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:584, 1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
 F:948-949/Cleavage site: Gly-Ile (collagenase) #status experimental
 F:1106/Binding site: carboxylate (Lys) (covalent) #status predicted
 Query Match 51.1%; Score 2370; DB 1; Length 1466;
 Best Local Similarity 49.5%; Pred. No. 4.9e-111;
 Matches 476; Conservative 56; Mismatches 285; Indels 144; Gaps 16;
 QY 1 GPPGPTGLPGRGGGSGRGP-----GADGAGVAGPKGAPAGER 42
 DB 234 GRGRGRGLPGRGGGSGRGP-----GADGAGVAGPKGAPAGER 293
 QY 43 GSPGAGPKGSGRGGGSGRGP-----GADGAGVAGPKGAPAGER 96
 DB 294 GAPGMPGRGAPGRGRGP-----GLPGAAGARGDSGQGGPPPTAGTGGSPGAK 350
 QY 97 ---GPPPPGARGGAGVGMFFGPK---GAAGPKAGRGVPPGPPGAVGPA---GKDGEA 147
 DB 351 GEVGPAGSGNSGAPGQGRGEPGQGHAGAQGPFGGINSPPGKGGEMGAPGAPGLM 410
 QY 148 GAQPPGAPGAGRGGGGSGRGP-----GADGAGVAGPKGAPAGER 207
 DB 411 GARGPPGAGAGPARGGAGGEPGKAGKEPGRGEGAGIPGVGAKGEGKDGSP 470
 QY 208 GERPTGLPGRGGGSGRGP-----GADGAGVAGPKGAPAGER 267
 DB 471 GEPGANGLPAAAGRGAPGRGAPNGIPGKGPAGRGAPGAPGAPGAPGAPGAPGAP 530
 QY 268 GLPGAAGLTSPPGDPGTPGPPGAPGQDGRGPPGPPGARGGAGVGMFFGPKGAPGER 327
 DB 531 GPGMRGMPGSGGSGRGP-----GADGAGVAGPKGAPAGER 590
 QY 328 GKAGRGVPPGPPGAGVAGPKGAGAGAGGPPGPPGAPGAGRGGGGSGRGP-----GADGAGVAGPKGAPAGER 387
 DB 591 GKNGRGVPPGPPGAGVAGPKGAGAGAGGPPGPPGAPGAGRGGGGSGRGP-----GADGAGVAGPKGAPAGER 650
 QY 388 GEAGKPGEGVPGDLGAP-----GPGSGPAGEPGPT 417
 DB 651 GENGKPGEGVPGDLGAP-----GPGSGPAGEPGPT 710
 QY 418 GLPGRGGRGGGSGRGP-----GADGAGVAGPKGAPAGER 462
 DB 711 GPPGPPGAGTGLGMPGEGGGLGSPGKDGEPGADGVPKDGPPGPPGPPGPPGPP 770
 QY 463 GEAGRP---GEAGLPKAGLTSPPGDPGTPGPPGAPGQDGRGPPGPPGAPGAGV- 518
 DB 771 GPAGQPGDKGEGGAGPLPGIAGPRGSPGERGTPGPPGAPGPPGPPGPPGPPGPPG 830
 QY 519 ---MGFPKGAAGSPGKAGRGVPP-----PGAAGVAGPKGAGAG 558
 DB 831 GEKGGGPPGAGVAGPPGSGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 890
 QY 559 GPPGAPGAGRGGGGSGRGP-----GADGAGVAGPKGAPAGER 618
 DB 891 GPPGPPGAGTGLGMPGEGGGLGSPGKDGEPGADGVPKDGPPGPPGPPGPPGPP 950
 QY 619 GPTGL-----GPPGPPGSGRGP-----PGAAGVAGPKGAPAGER 666
 DB 951 GITGARGLAGPPGMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1010
 QY 667 GEAGRPGAGLPGA-----KGLTSGPSGPPGDKTGPDP-----AGQDGRPPFP 711

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:17:54 ; Search time 121.782 Seconds

(without alignments)
3452.202 Million cell updates/sec

Title: US-10-658-989A-3

Perfect score: 4640

Sequence: 1 GPPGPGPTGLPGPPGRRG.....GEGQVPGDLGAPSPGAGG 821

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	2765	59.6	1 C111 HUMAN	P02452 homo sapien
2	2761	59.5	2 O76045	O76045 homo sapien
3	2761	59.5	2 Q8N473	Q8N473 homo sapien
4	2759	59.5	1 C111 CANFA	Q9X8J7 canis fami
5	2748.5	59.2	2 Q63079	Q63079 rattus norv
6	2732.5	58.9	1 C111 MOUSE	P11087 mus musculu
7	2732.5	58.9	2 Q810J9	Q810J9 mus musculu
8	2681.5	57.8	2 Q6LAN8	Q6LAN8 homo sapien
9	2679	57.7	1 C111 CHICK	P02457 gallus gall
10	2568	55.3	2 O93251	O93251 rana catesb
11	2561.5	55.2	2 O9YIB4	O9YIB4 cynops pyrr
12	2558.5	55.1	2 Q640B2	Q640B2 xenopus tro
13	2529	54.5	2 Q802B5	Q802B5 xenopus lae
14	2494	53.8	2 Q6PCL3	Q6PCL3 mus musculu
15	2486	53.6	2 Q14046	Q14046 homo sapien
16	2486	53.6	2 Q14047	Q14047 homo sapien
17	2481	53.5	2 Q77753	Q77753 canis fami
18	2479	53.4	2 Q91B91	Q91B91 xenopus lae
19	2477	53.4	2 Q28396	Q28396 equus cabal
20	2476	53.4	1 C112 HUMAN	P02458 homo sapien
21	2475.5	53.4	2 Q90W37	Q90W37 gallus gall
22	2465.5	53.1	2 Q77227	Q77227 gallus gall
23	2464	53.1	2 Q62031	Q62031 mus musculu
24	2464	53.1	2 Q62033	Q62033 mus musculu
25	2464	53.1	1 C112 MOUSE	P28481 mus musculu
26	2464	53.1	1 Q62032	Q62032 mus musculu
27	2455	52.9	2 Q80VY3	Q80VY3 mus musculu
28	2455	52.9	2 Q641K3	Q641K3 mus musculu
29	2454	52.9	2 Q63123	Q63123 rattus norv
30	2449	52.8	2 Q80X38	Q80X38 mus musculu
31	2430	52.4	2 Q6P4U1	Q6P4U1 brachydanio

32	2428.5	52.3	1447	2	Q6ULJ5	Q6ulj5 brachydanio
33	2412	52.0	1418	2	Q9W7E9	Q9w7e9 cynops pyrr
34	2405	51.8	1492	2	Q6P4Z2	Q6p4z2 xenopus tro
35	2401	51.7	1486	2	Q91717	Q91717 xenopus lae
36	2399	51.7	1486	2	Q7ZT16	Q7zt16 xenopus lae
37	2390	51.5	1449	2	Q6PEI9	Q6pei9 brachydanio
38	2386	51.4	1491	2	Q91718	Q91718 xenopus lae
39	2385	51.4	1491	2	Q7ZTM4	Q7ztm4 xenopus lae
40	2384	51.4	1449	2	Q6NZ15	Q6nzt5 brachydanio
41	2379	51.3	1049	1	CA13_BOVIN	P04258 bos taurus
42	2370	51.1	1466	1	CA13_HUMAN	P08121 mus sapien
43	2368.5	51.0	1464	1	CA13_MOUSE	Q7tt32 mus musculu
44	2368.5	51.0	1464	2	Q7TT32	Q7tt32 mus musculu
45	2368.5	51.0	1464	2	Q8BKX2	Q8bky2 mus musculu

ALIGNMENTS

RESULT 1

ID	CALL_HUMAN	STANDARD;	PRT;	1464 AA.
AC	P02452; P78441; Q13896; Q13902; Q13903; Q14037; Q14992; Q15176; Q15201; Q16050; Q7KZ30; Q7KZ34; Q8IVI5; Q9UML6; Q9UMW7;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	25-JAN-2005 (Rel. 46, Last annotation update)			
DE	Collagen alpha 1(I) chain precursor.			
GN	Name=COL1A1;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND VARIANT.SER-1434.			
RA	Dalgleish R.;			
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 1-589 FROM N.A.			
RX	MEDLINE=88329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;			
RA	D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,			
RA	Pretorius P.J.;			
RT	"Complete nucleotide sequence of the region encompassing the first			
RT	twenty-five exons of the human pro alpha 1(I) collagen gene			
RL	(COL1A1).";			
RN	[3]			
RP	SEQUENCE OF 1-472 FROM N.A.			
RX	MEDLINE=89025644; PubMed=3178743;			
RA	Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,			
RA	Jaanisch R., Prockup D.J.;			
RT	"Structure of a full-length cDNA clone for the prepro alpha 1(I) chain			
RT	of human type I procollagen.";			
RL	Biochem. J. 253:919-922(1988).			
RN	[4]			
RP	SEQUENCE OF 1-181 FROM N.A.			
RX	MEDLINE=84270697; PubMed=6462220;			
RA	Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,			
RA	Myers J., Williams C., Ramirez F.;			
RT	"Human pro alpha 1(I) collagen gene structure reveals evolutionary			
RT	conservation of a pattern of introns and exons.";			
RL	Nature 310:337-340(1984).			
RN	[5]			
RP	SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.			
RC	TISSUE=Skin;			
RX	MEDLINE=71038625; PubMed=5529814;			
RA	Click E.M., Bornstein P.;			
RT	"Isolation and characterization of the cyanogen bromide peptides from			
RT	the alpha 1 and alpha 2 chains of human skin collagen.";			
RL	Biochemistry 9:4699-4706(1970).			
RN	[6]			
RP	SEQUENCE OF 425-1464 FROM N.A.			
RX	MEDLINE=84080385; PubMed=6689127;			

RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
 RA Prockop D.J.;
 RA "Nucleotide sequences of complementary deoxyribonucleic acids for the
 RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
 RT of structures that are conserved during evolution.";
 RL Biochemistry 22:5213-5223(1983).
 RN [7]
 RN SEQUENCE OF 472-607 FROM N.A.
 RP PubMed=2981843;
 RX Chu M.-L., Gargiulo V., Williams C.J., Ramirez F.;
 RA "Multiclonal deletion in an osteogenesis imperfecta variant with
 RT increased type III collagen mRNA.";
 RL J. Biol. Chem. 260:691-694(1985).
 RN [8]
 RN SEQUENCE OF 488-625 FROM N.A.
 RP PubMed=3857621;
 RX Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gelinas R.E.;
 RA "Intron-mediated recombination may cause a deletion in an alpha 1 type
 RT I collagen chain in a lethal form of osteogenesis imperfecta.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874(1985).
 RN [9]
 RN SEQUENCE OF 1179-1464 FROM N.A.; VARIANTS OI-II HIS-1277; ARG-1388 AND
 RP 1337-GLU-TYR-1338 DEL, AND VARIANT SER-1434.
 RX MEDLINE=93352646; PubMed=8349697;
 RA Chessler S.D., Wallis G.A., Byers P.H.;
 RT "Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I)
 RT chain of type I collagen result in defective chain association and
 RT produce lethal osteogenesis imperfecta.";
 RL J. Biol. Chem. 268:18218-18225(1993).
 RN [10]
 RN SEQUENCE OF 1229-1454 FROM N.A.
 RP TISSUE=Bone;
 RX MEDLINE=88124208; PubMed=3340531;
 RA Maekela J.K., Raassina M., Virta A., Vuorio E.;
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
 RT domain.";
 RL Nucleic Acids Res. 16:349-349(1988).
 RN [11]
 RN SEQUENCE OF 1-34 FROM N.A.
 RP MEDLINE=85130970; PubMed=2857713;
 RX Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;
 RA "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
 RT Promoter structure, AluI repeats, and polymorphic transcripts.";
 RL J. Biol. Chem. 260:2315-2320(1985).
 RN [12]
 RN SEQUENCE OF 1-34 FROM N.A.
 RP MEDLINE=88097389; PubMed=3480516;
 RX Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
 RA "Regulatory elements in the first intron contribute to transcriptional
 RT control of the human alpha 1(I) collagen gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
 RN [13]
 RN SEQUENCE OF 1-44 FROM N.A.
 RP MEDLINE=88033098; PubMed=2822714;
 RX Rousow C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
 RA de Wet W.;
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)
 RT collagen gene enhance transcription.";
 RL J. Biol. Chem. 262:15151-15157(1987).
 RN [14]
 RN SEQUENCE OF 33-52.
 RP PubMed=2318855;
 RX Wirtz M.K., Keene D.R., Hori H., Glanville R.W., Steinmann B.,
 RA Rao V.H., Hollister D.W.;
 RT "In vivo and in vitro noncovalent association of excised alpha 1 (I)
 RT amino-terminal propeptides with mutant pro alpha 2(I) collagen chains
 RT in native mutant collagen in a case of Ehlers-Danlos syndrome, type
 RT VII.";
 RL J. Biol. Chem. 265:6312-6317(1990).
 RN [15]
 RN SEQUENCE OF 156-183 FROM N.A.
 RP PubMed=2767050;
 RX Weil D., D'Alessio M., Ramirez F., de Wet W., Cole W.G., Chan D.,
 RA Maatta A., Bornstein P., Penttinen R.P.;

RA Bateman J.F.;
 RT "A base substitution in the exon of a collagen gene causes alternative
 RT splicing and generates a structurally abnormal polypeptide in a
 RT patient with Ehlers-Danlos syndrome type VII.";
 RL EMBO J. 8:1705-1710(1989).
 RN [16]
 RN SEQUENCE OF 175-187 AND 274-289.
 RP PubMed=2169412;
 RX Baetge B., Notbohm H., Diebold J., Lehmann H., Bodo M., Deutzmann R.,
 RA Muller P.K.;
 RT "A critical crosslink region in human-bone-derived collagen type I.
 RT Specific cleavage site at residue Leu95.";
 RL Eur. J. Biochem. 192:153-159(1990).
 RN [17]
 RN SEQUENCE OF 263-268.
 RP TISSUE=Skin;
 RX MEDLINE=71001508; PubMed=4319110;
 RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
 RT "A comparative study of glycopeptides derived from selected vertebrate
 RT collagens. A possible role of the carbohydrate in fibril formation.";
 RL J. Biol. Chem. 245:5042-5048(1970).
 RN [18]
 RN SEQUENCE OF 281-302; 402-420; 823-842; 924-944; 1026-1045 AND
 RP 1143-1162 FROM N.A.
 RX PubMed=2374517;
 RA Labhard M.E., Hollister D.W.;
 RT "Segmental amplification of the entire helical and telopeptide regions
 RT of the cDNA for human alpha 1 (I) collagen.";
 RL Matrix 10:124-130(1990).
 RN [19]
 RN SEQUENCE OF 425-490; 965-1024; 999-1039 AND 1453-1464 FROM N.A.
 RP MEDLINE=83064528; PubMed=6183642;
 RX Chu M.-L., Myers J.C., Bernard M.P., Ding J.-F., Ramirez F.;
 RA "Cloning and characterization of five overlapping cDNAs specific for
 RT the human pro alpha 1(I) collagen chain.";
 RL Nucleic Acids Res. 10:5925-5934(1982).
 RN [20]
 RN SEQUENCE OF 710-745 FROM N.A., AND VARIANT OI-II ARG-728.
 RP PubMed=2339700;
 RX Wallis G.A., Starman B.J., Zinn A.B., Byers P.H.;
 RA "Variable expression of osteogenesis imperfecta in a nuclear family is
 RT explained by somatic mosaicism for a lethal point mutation in the
 RT alpha 1(I) gene (COL1A1) of type I collagen in a parent.";
 RL Am. J. Hum. Genet. 46:1034-1040(1990).
 RN [21]
 RN SEQUENCE OF 746-781 FROM N.A., AND VARIANT OI-III SER-767.
 RP MEDLINE=95187161; PubMed=7881420;
 RX Forlino A., Zolezzi F., Valli M., Pignatti P.F., Cetta G.,
 RA Brunelli P.C., Mottes M.;
 RT "Severe (type III) osteogenesis imperfecta due to glycine
 RT substitutions in the central domain of the collagen triple helix.";
 RL Hum. Mol. Genet. 3:2201-2206(1994).
 RN [22]
 RN SEQUENCE OF 1187-1220 FROM N.A., AND VARIANT CYS-1195.
 RP PubMed=3170557;
 RX Conn D.H., Apone S., Eyre D.R., Starman B.J., Andreasen P.,
 RA Charbonneau H., Nicholls A.C., Pope F.M., Byers P.H.;
 RA "Substitution of cysteine for glycine within the carboxyl-terminal
 RT telopeptide of the alpha 1 chain of type I collagen produces mild
 RT osteogenesis imperfecta.";
 RL J. Biol. Chem. 263:14605-14607(1988).
 RN [23]
 RN SEQUENCE OF 1440-1464 FROM N.A.
 RP MEDLINE=90110490; PubMed=2295701;
 RX Walling M.C., Cohn D.H., Byers P.H.;
 RA "Frameshift mutation near the 3' end of the COL1A1 gene of type I
 RT collagen predicts an elongated pro alpha 1(I) chain and results in
 RT osteogenesis imperfecta type I.";
 RL J. Clin. Invest. 85:282-290(1990).
 RN [24]
 RN SEQUENCE OF 1454-1464 FROM N.A.
 RP MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
 RX Maatta A., Bornstein P., Penttinen R.P.;

RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";

Query Match 59.6%; Score 2765; DB 1; Length 1464;
Best Local Similarity 55.4%; Pred. No. 8.2e-103;
Matches 561; Conservative 31; Mismatches 228; Indels 192; Gaps 19;
QY 1 GPPGEGPTGLPGRGGRGSGRFGADGAGPKGAGGSGGPGAGPKGSGEA--- 57
Db 179 GPMGSGRGLPGPPGAGPGQFGGPEGPGASGPMGPRGPPGPKNGDDGEAGK 238
QY 58 GRPGEAGLP---GAKGLTSGSGSPGPP-----DGKTGPPGAGODRPGPPGPGAR 105
Db 239 GRPGERGPPGQAGRLGPTAGLFCMKHGRFSGLDGAKGAGAGPKGEGSGENGAP 298
QY 106 GQAGVMGPPGPKGAAGEPGKAGER-----GVPPGPGAVGPA-----GKDGEAGAG 150
Db 299 GQMGPRGLPGRGRPGAGPAGARGNDGATCAAGPPGPGTGPAGPPGPGAVGAKGEAGP 358
QY 151 GPPGAGPAGERGPGP-----AGSGPFGQLPGPA 180
Db 359 GPRSEGGVGRGPPGPPGPPGAGAGPAGNPGADQPGKAGANGAPGTAGAPGFPGAR 418
QY 181 GPPGEAGKPGGQGVPGDLGAP-----GPSGPA-----GEPGPT 213
Db 419 GPGGPPGPPGKNSGEFCAFGSKGDTCAKGEPPGVQPPGAGEKKGARGEPGPT 478
QY 214 GLPDPGGRGPPGSRGFFGADGAGVAGPKGAGERSGPGAGKSGFGEAGRGEGAGLPGAK 273
Db 479 GLPDPGGRGPPGSRGFFGADGAGVAGPKGAGERSGPGAGKSGFGEAGRGEGAGLPGAK 538
QY 274 GLTSGPSGPDGKTGPPGAGODRPPGPPGARGOAGVMGPPGKGAERGPKAGER 333
Db 539 GLTSGPSGPDGKTGPPGAGODRPPGPPGARGOAGVMGPPGKGAERGPKAGER 598
QY 334 GVPGPAGVAGPAGKDEAGAGGPPGAGPAGEREGQAGSPGFGQLPGPAGPGEAGK 393
Db 599 GVPGPAGVAGPAGKDEAGAGGPPGAGPAGEREGQAGSPGFGQLPGPAGPGEAGK 658
QY 394 GEQGVPGDLGAPGSGPAGE-----PGP-----TGLPGPP 423
Db 659 GEQGVPGDLGAPGSGPAGE-----PGP-----TGLPGPP 478
QY 424 GERGGPSRGPAGDAGVAGPKGAGERSGPGAGKSGFGEAGRGEGAGLPGAK 474
Db 719 GSQAGPLQGMGPERGAAGLPGKGDGADGPKGADGSPGKDGVRGLTGPICPPGAGAP 778
QY 475 GAK-----GLTSGPSGPDGKTGPPGAGODRPPGPPGARGO----- 515
Db 779 GDKESGSGPAGTGAAGAPGDRGEPGPPGAGPAGPPGADGQKAGEPGDAGAKGDA 838
QY 516 -----AGVMGPPGPKGAAGEPGKAGERGVPPGPA-----VGPAGKDGAGAGAG 561
Db 839 GPPGAGPAGPPIGNTGAGCAKAGARSAGPPGATGFPAGAGRVGPPGSGNAGPPGPP 898
QY 562 GPAGP-----AGERGEQPGSGPFGQLPGPAGPAGPAGKPGGQGVPGDLAGPSPGAGE 618
Db 899 GPAGKEGKGRGTGTPAGRGVPPGPPGAGEKSGFADGAPAGATGTPGPGQIAGOR 958
QY 619 GPTGLPGRGGRGPPGSRGFFGADGAGVAGPKGAGERSGPPAGP-----KSGPCEAGRPGEA 675
Db 959 GVVGLPFGRGGRGFFGLPGSPGPKQKPGSGASGERGPPGMPGPPGLAGPFGESREGAP 1018
QY 676 GLPGAKGLTSGPSGPPGDKTGPAGODRPPGPPGARGOAGVMGPPGPKGAAGEP 735
Db 1019 GABGSGRGGSPGAKGDRGTGTPAGPAGPAGPAGPAGVPGAKSGDRGTGTPAGPAGPV 1078
QY 736 GKAGRGVPPGPPGAGVAGPAGKAGE-----AGAGGPPGP-----AGPA 771
Db 1079 GPAGARGPAGPQGRGDKGTGCTGEGQDRIKGRGSGGLQGPAGPPGSGPQEGPSGASGPA 1138
QY 772 GERGGQGPAGSP-----GFGQLPGPAGPAGPAGKPGGQGVPGDLAGPSPGAGP 820

Db 1139 GPRPPGSGAGAGKGLGLGPGPIGPPGRGTGAGVGPVGPVGPVGPVGPVGPVGP 1190
RESULT 2
O76045 PRELIMINARY; PRT; 1461 AA.
AC O76045;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-2004 (TREMBlrel. 26, Last annotation update)
DE Pro alpha 1(I) collagen.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.L., de Wet W., Bernard M., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, Alu repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
RA D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene.";
RL Gene 67:105-115(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
RA Maatta A., Bornstein P., Penttinen R.P.;
RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";
RL FEBS Lett. 279:9-13(1991).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=92157916; PubMed=1787829;
RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA Olsen A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL Matrix 11:375-379(1991).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=98107942; PubMed=9443882;
RA Korkko J., Ala-Kokko L., De Paeppe A., Nuytink L., Earley J.,
RA Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
RT scanning by conformation-sensitive gel electrophoresis identifies only
RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I;
RT Identification of common sequences of null-allele mutations.";
RL Am. J. Hum. Genet. 62:98-110(1998).
RN [7]
RP SEQUENCE FROM N.A.
RA Korkko J.M., Earley J.J., Nuytink L., DePaeppe A., Prockop D.J.,
RA Ala-Kokko L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF017178; AAB94054.2; -;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.

```
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg helix; 3.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS01184; VWF_C_2; 1.
KW Collagen.
SQ SEQUENCE 1461 AA; 138629 MW; 9ACF6DE30EA78E21 CRC64;

Query Match 59.5%; Score 2761; DB 2; Length 1461;
Best Local Similarity 55.3%; Pred. No. 1.2e-102;
Matches 560; Conservative 31; Mismatches 229; Indels 192; Gaps 19;

QY 1 GPPGEPGPTGLPGPPGRRGGGSRGFFGADGVAGPKGAGRGSPGAPGKSPGGEA--- 57
DB 176 GPMGSPGPRGLPGPPGAPGQGFQPPGEPGEGASGPMGPRGPPGPKNGDDGAGKP 235
QY 58 GRGEGALP---GAKGLTSGSPGSP-----DKTQPPGAPGQDRPGPPPGAR 105
DB 236 GRGEPGPPGQAGRLPGTAGLPGMKHGRGFSGLDGAKGDAGPAGPKGEPSPGNGAP 295
QY 106 GQAGVMGFPKGAAGBPGKAGER-----GVPPGAVGPA-----GKDGAGAQ 150
DB 296 GQMGPRGLPGRGPRGAPGAGRGNDGATGAAGPFGTGPAGPPGPPGAVGAKGAGPQ 355
QY 151 GPPGAPGAGERGQGP-----AGSPGFGQLPGPA 180
DB 356 GPRGSEPGQVRGEPGPPGAGAGPAGNPDGQPGKANGANGAPGAGPFGGARGPS 415
QY 181 GPPGEGAKPGEQVPGDLGAP-----GPSGA-----GEPGPT 213
DB 416 GPQGPGGPPGPKNSGEPGSGKDTGAKGEPGVGVQGPFGPAGGEGRGARGEPGPT 475
QY 214 GLPGPPGERGSGSRGFPFGADGVAGPKGAGRGSPGAPGKSGPCEAGRPGGAGLPKAK 273
DB 476 GLPGPPGERGSGSRGFPFGADGVAGPKGAGRGSPGAPGKSGPCEAGRPGGAGLPKAK 535
QY 274 GLTSGSPGPDGKTGTPGAGQDGRPPGPPGARGQAGVMGFPKGAAGEPGKAGER 333
DB 536 GLTSGSPGPDGKTGTPGAGQDGRPPGPPGARGQAGVMGFPKGAAGEPGKAGER 595
QY 334 GVPGPPGAVGAPKDGEGAGAGQPPGAPGAGRGEGQPGAGSPGFGQLPGPAGPGEAGKP 393
DB 596 GVPGPPGAVGAPKDGEGAGAGQPPGAPGAGRGEGQPGAGSPGFGQLPGPAGPGEAGKP 655
QY 394 GEGVPGDLGAPGSPGAGE-----PGP-----TGLPGPP 423
DB 656 GEGVPGDLGAPGSPGARGERGPPGPPGAPGPRGANGAPGNDCAKADGAPGAP 715
QY 424 GERGGPSRGFPAGDGVAGPKGAPGAGRGSPGAPGKSGPGE-----AGRPGGAGLP 474
DB 716 GSQAGFLQGMPEGGAAGLPKGDGAGPKGADGSPGKDGVRGLTGPFGPPGAPGAP 775
QY 475 GAK-----GLTSGSPGPDGKTGTPGAGQDGRPPGPPGARGQ----- 515
DB 776 GDKGESGSPGAPGTGARGAGDGRGFPFGPPGAGFAGPGGADGQPGKAGEPGDAGAKGDA 835
QY 516 -----AGVMGFPKGAAGEPGKAGERGVGPPGPA-----VGPAGKDGEGAGAGPP 561
DB 836 GPPGAPGAPGPPGPIGNVGPAGKAGARGAGSAGFPFGATGPPGAGRVGPPGSPGNAAGPPPP 895
QY 562 GPAGP---AGERGEGQPGSPGFGQLPGPAGPPGAGKPGEGGVGPDGLGAPGSPGAGEP 618
DB 896 GPAGKGGGKPGRTGTGAPRGPEVGPFGPPGAPGKSGPGADGAPGAPGTPGPGQIAGQR 955
QY 619 GPTGLPFPGRGGGSGRFPFGADGVAGPKGAPGAGRGSPGAPG---KGSPGEGAGRPGEA 675
```

```
Db 956 GVVGLPGQGRGERGFGPLPGSPGEPCKQKQSGASGERGPPGPGPLAGPFGSGREGAP 1015
QY 676 GLPGAKGLTSGSPGPDGKTOPPAGQDGRPPGPPGARGQAGVMGFPKGAAGEP 735
Db 1016 GAEGSPGRDGSFGAKGDRGETGTPAGPAGPAGPVPVGPAGKSGDRGETGTGAPGAPV 1075
QY 736 GKAGRGVPGPPGAVGAPGAKDGE-----AGAGGPPGP-----AGPA 771
Db 1076 GPPGARGPAGPQGRDGETGEGQDRIKGRGFGSLGPPGPPGSGCEGSPGSGASGPA 1135
QY 772 GERGQGPAGSP---GFGQLPGPAGPGEAGKPGEQGVGDLGAPGSPGAP 820
Db 1136 GPRGPPGSAGAPGKGLNGLPGPIGPPGPRGTGAGPVGPPGPPGPPG 1187

RESULT 3
Q8N473 PRELIMINARY; PRT; 1464 AA.
AC Q8N473;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-WAR-2004 (TRENBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen, preproprotein.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., McQuellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., Lebowitz P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Wolzky K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX Strausberg R.;
RA Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC036531; AAH36531.1;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR01007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg helix; 3.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00038; COLFI; 1.
```



```
FT MOD_RES 261 261 5-hydroxylysine (By similarity).
FT MOD_RES 1160 1160 3-hydroxyproline (By similarity).
FT CARBOHYD 261 261 O-linked (Gal... ) (By similarity).
FT CARBOHYD 1361 1361 N-linked (GlcNAc... ) (By similarity).
FT VARIANT 208 208 G -> A (in OI; severe).
SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;

Query Match 59.5%; Score 2759; DB 1; Length 1460;
Best Local Similarity 55.9%; Pred. No. 1.4e-102;
Matches 559; Conservative 31; Mismatches 230; Indels 180; Gaps 19;

QY 1 GPGEGPTGLPGPPGGERGPGS-----RGFPCA-----DGVAGPKPAGERSGPGA 48
Db 187 GPGAGPQFQPGPGEFGASGPMGRPPGPKNGKDDGEAKGPKRGERGPPGQ 246
QY 49 GPKGSFGEAGRP-----GEAGLPKAGKLTGSPGSPGDPDKTTPPGAGQD 93
Db 247 GARGLPGTAGLPMKMGHGFSGLDGAKDAGPAGPKGSPGSENGAPQMGRLPGER 306
QY 94 GRPGPPGPPGARGQ---AGVMGPPGKGAAGE---FGKAGERGVPPGCAVGPAGKGEA 147
Db 307 GRPGAGPAGARGNDGATGAAGPPGPTGAGPPGFGCAVCAKGEAGPQARGSEGPQYR 366
QY 148 GAQGPPGPPGAGERGEQGP-----AGSPGFQGLPGPAGPPGGEAKGPEQ 192
Db 367 GEPGPPGPPGAGPAGNPGADQPGKAGKANGAPGAGPFGPARGPSPGQSPGPK 426
QY 193 GVPGLDGLAP-----GPGSPA-----GEPGPTGLPPPGGERGPP 225
Db 427 GNSGEPGAPGNKDTGAKGPPGTGLOGPPGAGEGKRGAGEPQGTGLPFPGERGPP 486
QY 226 GSRGFPAGDVGAPKGPAGERSGPPGAPKGSFGEAGRPCEAGLPKAKGLTSGSPGPD 285
Db 487 GSRGFPAGDVGAPKGPAGERSGPPGAPKGSFGEAGRPCEAGLPKAKGLTSGSPGPD 546
QY 286 GKTGPPGPPAGDGRPPGPPGARGQAGVMGPPGPKAAGEPKKAGERGVPPGCAVGA 345
Db 547 GKTGPPGPPAGDGRPPGPPGARGQAGVMGPPGPKAAGEPKKAGERGVPPGCAVGA 606
QY 346 GKDGEGAAGQPPGAPGAGERGQGPAGSPGFQGLPGPAGPPGCEAGKPGEQGVPGDLGAP 405
Db 607 GKDGEGAAGQPPGAPGAGERGQGPAGSPGFQGLPGPAGPPGCEAGKPGEQGVPGDLGAP 666
QY 406 GPSGPPAGE-----PGP-----TGLPFPGERGPPGSRGPP 435
Db 667 GPSGAGERGPPGERGVQGPAGPAGPRGANGAPNDGAKGDAGAPGAPGSGAPLQGM 726
QY 436 GADGVAGPKPAGERGSGPPGAPKGSFGE-----AGRPGEAGLPKAK----- 477
Db 727 GERGAAGLPKPKDGRDADGPKGADGSPGKDGVRGLTGPIGPPGAPAGDCKGEAGSPGA 786
QY 478 GLTSGSPGPPGDKTQPPGAPGQDGRPPGPPGPPGARGQ-----AGVMGPP 522
Db 787 GPTGARGAPDRGEPGPPGAPGAPGPPGADGQPGAKGEPGDAGKADGAPPGPTGPP 846
QY 523 GPKGAAGEPKKAGERGVPPGPA-----VGPAGKDGEGAQAGPPGAPG----- 570
Db 847 GPTGNVAGPPKAGRSAGSPGATGPPGAAGRVGPPGSGNAGPVPGPPGAKGEGKGAR 906
QY 571 GEOGPAGSPGFQGLPGPAGPCEAGKPGEQGVPGDLGAPGSPGAGEPPTGLPGPGER 630
Db 907 GETGPAGRGVGVPPGPPGAPGKSGFADGAGAPGTPGPGIAGQRRVGLPQQRGER 966
QY 631 QGSPGSGFFGADVAGPKPAGERSGPPGAPG---KGSFGEAGRPCEAGLPKAKGLTSP 687
Db 967 GPFPLPGSPGEPGKQGSFSGTSGERGPPGPPGPPGLAGPPCESGREGSPGAESSPGRDGSP 1026
QY 688 GSPGPPGKTPGPPAGQDGRPPGPPGARGQAGVMGPPGPPGKAGGEPKAGERGVPPPP 747
Db 1027 GPKGDRGETGAPGPPGAPGAPGPPGPPGARGVGPAGKNGDRGETGPPAGPAGPIG 1086
QY 748 GAVGPAKDKGE-----AGAQQPPGP-----AGPAGERGEQCPAGSP 783
```

```
Db 1087 GPRGDKGTGEGQDRGIRKGRHGFSGLOGPPGPPGSGQPSGASGAPAGPPGSGAGSP 1146
QY 784 ---GFQGLPGPAGPPGEGAKGPEQGVPGDLGAPGSPGAG 820
Db 1147 GKDLGNLPGPIGPPGPRGRTGDAGVGPVPPGPPGPPGPPG 1186

RESULT 5
Q63079 PRELIMINARY; PRT; 1453 AA.
AC Q63079;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Collagen alpha1 (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RX SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Bone, and Tooth;
RX MEDLINE=99163824; PubMed=10065941;
RA Brandsten C., Lundmark C., Christersson C., Hammarstrom L., Wurtz T.;
RT "Expression of collagen alpha1(I) mRNA variants during tooth and bone
formation in the rat.";
RL J. Dent. Res. 78:11-19(1999).
DR EMBL; Z78279; CAB01633.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clq helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib-collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR010007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clq helix; 3.
DR ProDom; PD002078; Fib-collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C.1; 1.
DR PROSITE; PS0184; VWF_C.2; 1.
KW Collagen.
FT NON TER
SQ SEQUENCE 1453 AA; 137886 MW; E6896BDC19A4A1D8 CRC64;

Query Match 59.2%; Score 2748.5; DB 2; Length 1453;
Best Local Similarity 52.3%; Pred. No. 3.7e-102;
Matches 556; Conservative 33; Mismatches 231; Indels 243; Gaps 19;

QY 1 GPGCE---PGTGLPGPPGERGPGSGRFPG-----ADGVAGPKPAGER 42
Db 117 GPPQDGLPGQGLPGPPGPPGPPGPPGLGLGNFASQMSYGYDEKSAGSVSPGPMGSPGR 176
QY 43 GSPGPAGPKGSPGAGRPFGEAGLPKAGLTCSPGSPGDKTGPAGAGQDGRP---Gpp 99
Db 177 GLPFPAGAPGQGGPFGPFGEGGSGPMGPPFPFPFGKNGDDGAGKFCRGERGPP 236
QY 100 GPPGARGQAGVMGPP-----GPKGAAGFPK-----A 126
Db 237 GPQAGARGLPGTAGLPGMKHGRFSGLDGAKGDTGAPGKGPSPGSENGTGPQMGPRGLP 296
QY 127 GERGVGPPGAVGAGKDGEGAQGPDP-----AGPAGERGEQGP- 167
Db 297 GERGRPPGPTAGARGNDGAVGAAGPPGPTGPTGPPGPAAGAKGAGPQCARSGEGPQ 356
QY 168 -----AGSPGFPQGLPGPAGPPGEGACKP 189
Db 357 GVRGPPGPPGAGAGPAGNPGADGQPCAKGANGAPGAGPFGARGPSPGPPSGAP 416
```


DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWC.1; 1.
 DR PROSITE; PS0184; VWC.2; 1.
 KW Collagen; Extracellular matrix; Glycoprotein; Hydroxylation;
 KW Pyrrolidone carboxylic acid; Repeat; Signal; Structural protein.
 FT SIGNAL 1 22
 FT PROPEP 23 151
 FT CHAIN 152 1207
 FT PROPEP 1208 1453
 FT DOMAIN 29 87
 FT DOMAIN 152 167
 FT DOMAIN 168 181
 FT DOMAIN 182 1207
 FT MOD_RES 152 152
 FT MOD_RES 160 160
 FT MOD_RES 254 254
 FT MOD_RES 1153 1153
 FT CARBOHYD 56 56
 FT CARBOHYD 254 254
 FT CARBOHYD 1354 1354
 FT SITE 734 736
 FT SITE 1082 1084
 FT CONFLICT 1450 1450
 FT SEQUENCE 1453 AA; 137944 MW; 3802E535DF81808 CRC64;
 Query Match 58.9%; Score 2732.5; DB 1; Length 1453;
 Best Local Similarity 54.7%; Pred. No. 1.6e-101;
 Matches 547; Conservative 33; Mismatches 240; Indels 180; Gaps 16;
 1 GPGGSGEAGRP-----RGSGRFFPGADGVAGPKGAGRGSPGPA 48
 180 GPPGAPGQGGPPGPFPGSGGPMRPGPPGPKGNDGAGKGRGPPGPGQ 239
 49 GPKGSGEAGRP-----GEAGLPGKGLTSGSGSPGPDGKTGPPGAGQD 93
 240 GARGLPCTAGLPGMKHGRGSLDGAKGADGAPGPKGSPGNGAPQMGRLPGR 299
 94 GRPGPPPPGARGQAGVWGPFGPKGAAGE-----FKAGRGVPPGPGVAGVAGPKDGEA 147
 300 GRFGPPGTAGAGNDGAVGAAGPPGTGTPGPFPGVAGVAGKGEAGPGQARGSEGPQYR 359
 148 GAQGPFGPAGPAGRGEPQ-----AGSGFQGLPGAPGPPGAEAGKPGQ 192
 360 GEPGPPGAGAGPAGNPGADGQPGKANGAGPAGIAGAPGFGARGSPGSPGPPGPK 419
 193 GVPFCDLGA-----GPGSGA-----GEPPTGLPGPPGPPGPPG 225
 420 GNSGEPGAPGKGDGTGAKGSPGATGVQPPGPGAGEGKRGARCEPFGSLPFPGERGGP 479
 226 GSRGFPAGDGVAGPKGPPGARGQAGVWGPFGPKGAAGE-----FKAGRGVPPGPGVAGVAGPKDGEA 147
 480 GSRGFPAGDGVAGPKGPPGARGQAGVWGPFGPKGAAGE-----FKAGRGVPPGPGVAGVAGPKDGEA 147
 286 GKTGPPGAGDQGRPPGPPGARGQAGVWGPFGPKGAAGE-----FKAGRGVPPGPGVAGVAGPKDGEA 147
 540 GKTGPPGAGDQGRPPGPPGARGQAGVWGPFGPKGAAGE-----FKAGRGVPPGPGVAGVAGPKDGEA 147
 346 GKDGAGAGQPPGAGPAGRGEPGPGFQGLPGAPGPPGAEAGKPGQVPPGDLGAP 405
 600 GKDGAGAGQPPGAGPAGRGEPGPGFQGLPGAPGPPGAEAGKPGQVPPGDLGAP 405
 406 GPGSGPAGE-----PGP-----TGLPGPPGPPGPPGPPGPPG 659
 660 GPGSARGRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 719
 436 GADGVAGPKGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 719
 720 GERGAAGLPKGRDAGPKGADGSPGKDGARGLTGPPGPPGAGAGDGGKAGSPGPP 779
 487 GPD-----GKTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 779
 780 GPTGARGAGDGRGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 780

QY 523 GPKGAAGBPKGAGRGVPPGPG-----VGPAGKDGAGAGQPPGPPGAGP-----AGER 570
 DB 840 GFIGNVAGFPKPGPRGAGPAGTGPFGAAGRVGPPGPGSGNAGPPGPPGPKGKGKGR 899
 QY 571 GQGGPAGSPGFOGLPGPAGPCEAGKPGQVPPDGLGARGSPGAGEPPTGLPFPGR 630
 DB 900 GTGTGAGRGVEVPPGPPGAGEKSPGADGAGSPGTPGPGIAGQGVVGLPQGRGR 959
 QY 631 GPGSRGPFAGDVGAGPKGPPGAGRGSPGAGP-----KSGPGEAGRGAGEAGLPGAKGLTGP 687
 DB 960 GFFGLPGSGEPGKQPSGSGRGPPGPMGPPGLAGPFGSGRGSGGSGAGSPGPDGAP 1019
 QY 688 GSPGPDGKTGPPGAGQGRPPGPPGARGQAGVWGPFGPKGAAGEPKGAGRGVPPG 746
 DB 1020 GAKGDRGETGPPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 1079
 QY 747 -----PGAVGPKGDEAGAGQPPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGP 783
 DB 1080 GPRGDKGTGQDGRGKIKGHRGFSGLQPPGSPGSGQSGSGAGSPGPPGPPGAGSP 1139
 QY 784 ---GFGQLPGPAGPGEAGKPGQVPPDGLGAPGSPGAG 820
 DB 1140 GKGGLGLPGTGPGRGTGDSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1179
 RESULT 7
 Q810J9 PRELIMINARY; PRT; 1453 AA.
 AC Q810J9
 ID Q810J9
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Procollagen, type I, alpha 1.
 GN Name=Colla1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Scheetz T.E.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalish U., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RX Strausberg R.;
 RA Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC050014; AAH50014.1;
 DR MGI:88467; Colliat.
 DR GO:0005615; C:extracellular space; TAS.
 DR InterPro; IPR008161; Clg helix.
 DR InterPro; IPR008160; Collagen.

DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; C1g helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS0184; VWF_C; 1.
KW Collagen.
SQ SEQUENCE 1453 AA; 138032 MW; 0B7F06BBB9A1D5EA CRC64;

Query Match 58.9%; Score 2732.5; DB 2; Length 1453;
Best Local Similarity 54.7%; Pred. No. 1.6e-101;
Matches 547; Conservative 33; Mismatches 240; Indels 180; Gaps 16;

QY 1 GPPGPGTGLPGPGE-----RGPGSRGPPGADGVAGKPGAGRGSPGPA 48
DB 180 GPPGAPGQGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 239
QY 49 GPKSGPGAGRP-----GBAGLPAGKGLTSGSPGPPGKTKGPPGPAQOD 93
DB 240 GARGLPGTAGLPGMKHGRGSLDGAAGDAGPAGKPGPPGPPGPPGPPGPPG 299
QY 94 GRPGPPGPPGAGQAGVGMGPPGPKAAGE-----PGKAGRGVPPGPPGAVG 147
DB 300 GRPGPPGTAGAGNDGAVGAGPPGTPGTGPPGPPGAVGAKGAGPPGARG 359
QY 148 GAQGGPPGPPGAGRGEGGP-----AGSPGQGLPGPPGPPGPPGPPGPP 192
DB 360 GEPGPPGPPGAGPAGNPGADQCPGKANGAPGIAGAPFPGARGSPGPPGPP 419
QY 193 GVPGLGAP-----GPSGA-----GEPGTLGPPGPPGPPGPPGPP 225
DB 420 GNSGPPGPPGKNDGTAKGPPGATGVQPPGPPGPPGPPGPPGPPGPPGPP 479
QY 226 SRRGPPGADGVAGKPPGAGRGSPGPPGPPGPPGPPGPPGPPGPPGPPGPP 285
DB 480 SRRGPPGADGVAGKPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 539
QY 286 GKTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 345
DB 540 GKTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 599
QY 346 GKDGAGAGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 405
DB 600 GKDGAGAGQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 659
QY 406 GPSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 435
DB 650 GPSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 719
QY 436 GADGVAGKPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 486
DB 720 GERGAAGLPKPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 779
QY 487 GPD-----GKTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 522
DB 780 GPTGARGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 839
QY 523 GPKGAAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 570
DB 840 GPTGNVAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 899
QY 571 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 630
DB 900 GTGTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 959
QY 631 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 687

DB 360 GFPLPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1019
QY 688 GSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 746
DB 1020 GAKGDRGTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1079
QY 747 -----PGAVGPPAGKDEAGAGQPPGPPGPPGPPGPPGPPGPP 783
DB 1080 GPRGDKGTGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1139
QY 784 ---GFQGLPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 820
DB 1140 GKDGGLGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1179

RESULT 8
Q6LAN8 PRELIMINARY; PRT; 1069 AA.

ID Q6LAN8
AC Q6LAN8;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Collagen type I alpha 1 (fragment).
GN Name-COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97141927; PubMed=8988177;
RA Simon M., Pedeutour F., O'Brien K., Sirvent N., Kedra D., Guilbaud C.,
RA Turc-Carel C., Dumanski J.P.;
RT "Regulation of the placental-derived growth factor B-chain gene via
RT fusion with collagen gene COL1A1 in dermatofibrosarcoma protuberans
RT and giant-cell fibroblastoma";
RL Nat. Genet. 15:95-98(1997).
RL [2]
RP SEQUENCE FROM N.A.
RA O'Brien K.P.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; X98705; CAA67261.1; -.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0006817; P-phosphate transport; IEA.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01391; Collagen; 16.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; C1g helix; 3.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS0184; VWF_C; 1.
KW Collagen.
FT NON TER 1069 1069
SQ SEQUENCE 1069 AA; 97445 MW; BE279B10572FB980 CRC64;

Query Match 57.8%; Score 2681.5; DB 2; Length 1069;
Best Local Similarity 58.4%; Pred. No. 1.4e-99;
Matches 524; Conservative 28; Mismatches 229; Indels 117; Gaps 10;

QY 1 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 60
DB 203 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 262
QY 61 GEAGLPKAGLITGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 120
DB 263 GMKGRGPPGLDAGKADGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 322
QY 121 GEPKAGRGVPPGPPGAVP-----AGKDEAGAGQPPGPPGPPGPPG 167
DB 323 GNDGATGAAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 382

DR ProDom: PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS0184; VWFC_2; 1.
KW Collagen; Direct protein sequencing; Extracellular matrix;
KW Glycoprotein; Hydroxylation; Pyrrolidone carboxylic acid; Repeat;
KW Signal; Structural protein.
FT SIGNAL 1 22
FT PROPEP 23 151 N-terminal propeptide.
FT CHAIN 152 1205 Collagen alpha 1(I) chain.
FT PROPEP 1206 1453 C-terminal propeptide.
FT DOMAIN 31 89 VWFC.
FT MOD_RES 152 152 Pyrrolidone carboxylic acid.
FT MOD_RES 160 160 Alanine (By similarity).
FT MOD_RES 254 254 5-hydroxylysine (By similarity).
FT MOD_RES 851 851 5-hydroxylysine (Potential).
FT MOD_RES 1081 1081 Hydroxyproline (Potential).
FT MOD_RES 1097 1097 5-hydroxylysine (Potential).
FT MOD_RES 1153 1153 3-hydroxyproline.
FT CARBOHYD 254 254 O-linked (Gal...) (By similarity).
FT CARBOHYD 1354 1354 N-linked (GlcNAc...) (By similarity).
FT CONFLICT 1187 1187 F -> L (in Ref. 5).
FT CONFLICT 1441 1441 Q -> H (in Ref. 6).
SQ SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;

Query Match 57.7%; Score 2679; DB 1; Length 1453;
Best Local Similarity 53.7%; Pred. No. 2.1e-99;
Matches 540; Conservative 30; Mismatches 250; Indels 186; Gaps 16;

QY 1 GPPGPGTGLPCPPGRRGSGRFPFGADGVAGPKGA-----GER 42
DB 174 GPRGLPGPPAGPQGFQPGPEGPGASGPMGPPGPPGKNGDDGEAGKPRPQQR 233
QY 43 GSPGAGPKGSGEAGRPCEAGLPCAKGLTSGSPGDKGTGPPGAGOD----- 93
DB 234 GPPGQAGRLPDTAGLPGMKHGFSLDGLAKQPGPAGPKGFGSGENGAPQCMQPR 293
QY 94 -----GRPGPPGPPCARGQ-----AGVMGPPGPKGAAGE-----PGKAGERGVPPGAVGPA 141
DB 294 GLPGERGRPPGSPAGARGNDGAPCAAGPPGTPGAGPPGPPGGAAGKGTGPGCARGSE 353
QY 142 GHDGAGAGCPGPPAGPAGERGEQPP-----AGSGFQGLPPGAPGPEA 186
DB 354 GPQSGRGEPPGPPGAGAGPAGNPGADGQPGAKGATGAPGIAGAPFPFGARPGSPGQPS 413
QY 187 GKPGQGVPPGDLGAP-----GPSGA-----GEPGTLGPP 219
DB 414 GAPGPKGNSGPPCAPGNKGDTCAGKPGGAGVQVPPGAGEGKRGAPGAPGLGPA 473
QY 220 GERGGSGRPPGADGVAGPKGAPAGERGSPGAPGKSPGGEAGRPCEAGLPCAKGLTSP 279
DB 474 GERGAPSGRPPGADGIAGPKPPGERSPGAVGPKSGFGEAGRPCEAGLPCAKGLTSP 533
QY 280 GSPGPDGTPGPPAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEPKAGERGVPPGP 339
DB 534 GSPGPDGTPGPPAGQDGRPPGPPGARGQAGVMGPPGPKGAAGEPKAGERGVPPGP 593
QY 340 GAVGAGKDGAGAGCPGPPAGPAGERGEPGQAGSPGFGGLPGAPGCEAGKPGCEQVP 399
DB 594 GAVGAGKDGAGAGCPGPPGTPGAGERGEPGQAGSPGFGGLPGAPGCEAGKPGCEQVP 653
QY 400 GDLGAPGSPGSPAGE-----PGP-----TGLPFPGERGP 429
DB 654 GNAGAPGAPAGERGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 713
QY 430 GSRGPPGADGVAGPKGAPAGERGSPG-----AGPKGSGEAGRP-----GEA 471
DB 714 GLEGMPPGERGAAGLPCAKGDRGDPGPKGADGAPKDGRLGLTGPFGPPGAPGPKGEA 773
QY 472 GLPGAKGLTSGSPGPPGDKTGPFGPA-----QDGRPPGP 507
DB 774 GPPGAGTGPAGAPGDRGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 833

QY 508 GPPGARGQAGVMGPPGPKGAAGEPKKAGERGVPPGPPGAVGAGKDGKGEAGAGPPGPPGPA 567
DB 834 GPTGAPGPPAGVAGPPGPKGARGSGAGPPGATGPPGAAGRVGPPGSGNIGLPGPPGPKGZ 893
QY 568 GE---RGSGQAGSPGFGGLPGPAGPPGCEAGKPGCEQGVPPGDLGAPGSPGAPGPPGPTGLP 624
DB 894 GSKPRGETGPPAGRPGEPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 953
QY 625 GPPGRRGSGRPPGADGVAGPKGAPAGERGSPGAPG---KGSFGEAGRPCEAGLPGAK 681
DB 954 GQRRGFPGLPGPSGBRGKQPSGASGGERGPPGPMGPPGLAGPPGEAGREGAPCAECAP 1013
QY 682 GLTSGPSGPPDGTGTPPPGAGQGRPCPPGPPGARGQAGVMGPPGPKGAAGEPKKAGER 741
DB 1014 GRDGAAGPKGRGETGPPAGPPGAPGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1073
QY 742 GVPGPPGAVGPPGKDG-----AGACQGPP-----GPAKAGER 774
DB 1074 GPAGPPGRRGKGTGEGQGRGMKGRHGFSLQPPGPPGPPGPPGPPGPPGPPGPPGPP 1133
QY 775 GEOQPPGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 820
DB 1134 GSAGAKGDLGLPGPIGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 1179

RESULT 10
O93251 PRELIMINARY; PRT; 1445 AA.

AC O93251;
DT 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen.
GN Name=alpha 1 type I collagen;
OS Rana catesbeiana (Bull. frog);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Rana;
OX NCBI_TaxID=8400;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9294154; PubMed=10367734; DOI=10.1016/S0945-053X(99)00005-0;
RA Asahina K., Utoh R., Obara M., Yoshizato K.;
RT "Cell-type specific and thyroid hormone-dependent expression of genes
of $\alpha 1(I)$ and $\alpha 2(I)$ collagen in intestine during
amphibian metamorphosis.";
RT Matrix Biol. 18:89-103(1999).
RL EMBL; AB015440; BAA29028.1;
DR GO; GO:0005581; C:cytoplasm; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Cig helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Cig_helix; 4.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS0184; VWFC_2; 1.
KW Collagen.
SQ SEQUENCE 1445 AA; 137251 MW; F59BB550C9873F04 CRC64;

Query Match 55.3%; Score 2568; DB 2; Length 1445;
Best Local Similarity 53.6%; Pred. No. 5.4e-95;
Matches 516; Conservative 43; Mismatches 259; Indels 144; Gaps 17;

QY 1 GPPGPGTGLPGPPGRRGSGRFPFGADGVAGPKGA-----RGFPGADGVAGPKGPA-----GER 42

```
Db 221 KXPGRGRRGPPGQAGLPGTAGLPGMKHGRFNGLDGAKGDTGTPAGKGEFGNGEN 280
QY 43 GSPGAPGKSGPGEAGRPGAGLPGKAGLGTGSPGSDGDKTGPDPGPPG-----A 90
Db 281 GAGQVQGRGLGGERGPPGSPAGARNDTTPGAGPFGTGTGTPGPPGPGVPGKGA 340
QY 91 GQDGRPPGPPGARGQAGVGMFFGPGKAAAGPCKAGRGVPPGPPGAGVPGAGKGEAG 150
Db 341 GPQSGRGGPQAGGEPGAPGQAGPAGSAGNPGTDGQPKAKGATGAPGIAGAPGPGAR 400
QY 151 GPPGAPGAGRGEGGAGSGPFGGLPGAPGPPGCEAGKPGCEQGVPGDLGAPGSGPAGEP 210
Db 401 GAGPQGGSGPCKGNNGEPGAGNKGEPGAKGESGAGSGQGGPPGPEGKRGSRGEP 460
QY 211 GPTGLPGPPGSGRGGPFGGADGVPAGKPGPAGRGSPGAPGKSGPGEAGRPGAGLP 270
Db 461 GPSGPPGAPGRGAPGSRGPPGADGAGKPGPPGGERGPPGAGKSGPGEAGRPGAGLP 520
QY 271 GAKGLTSGSPGSDGDKTTPGPPAGQDGRPPGPPGARGQAGVGMPPGPKGAAGPFGKA 330
Db 521 GAKGLTSGSPGSDGDKTTPGPPAGQDGRPPGPPGARGQAGVGMPPGPKGAAGPFGKP 580
QY 331 GERGVPPGAVGAGKDGAGAGCPGPPAGPAGRGEGGQPGSGPFGQLPGPAGPGEA 390
Db 581 GERGVAGPPGAVGAGKDGAGAGCPGPPAGPAGRGEGGQPGSGPFGQLPGPAGPGEA 640
QY 391 GKPGEGQVPGD-----LGAPGSGP-----AGBP 414
Db 641 GKPGEGQVPGDVGSPGAGSRGPPGGERGALPGPPGQFPGANGANGAKGAGAP 700
QY 415 -----GPTGLPGPPGSG-----GP-GSRGPPGADGV-----AGKPGAGRR 450
Db 701 GAGCGQGGPGLQMGPPGRRGAGLPGKAGDGRDQKPGADGAPKDGVRGLTGPFGPG 760
QY 451 GSP-----GPAGP---KSGPGEAGRPGAGLPGKAGLGTGSPGSDGDKTGPFGPA 498
Db 761 GAGDKGEGAGPAGPAGTGSRGAPGERGPPGSGPAGPAGPAGDQPKAGGQDAGBK 820
QY 499 QDGRGPPGPPGARGQAGVGMPPGPKGAAGPCKAGRGVPPGPPGAGVAGPAGKDGAG 558
Db 821 GDAGPPGAAAGTGAAGPAGAVGATGPKAGPAGPFGSTGFPGAAGRVGPPGSGNAGPP 880
QY 559 GPPGAPGAPAGE---RGEQGAGSPGQGLPGPAGPGEAGKPGCEQGVPGDLGAPGSGPA 615
Db 881 GPSGAPGKEQKQPRGETGAPRGPFAGAGPPGSGGKSGPDSGDPAGPAGIPGPGGTA 940
QY 616 GEPGPTGLPDPGRRGPPGPPGADGVAGPKPAGRGSPGAPGKSGPGEAGRPGEA 675
Db 941 GTRGTGVLPGQRRGPPGPGTGPFGKQSSGSPGSGRPPGPPGGLAGPPGPPGRE 1000
QY 676 GLPGAKGL-----TGSPGSDGDKTGPFGAGQD---GRPPGPPPP 714
Db 1001 GSPGSGSPGSDGAPKDGSGSPGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 1060
QY 715 GARGQAGVGMPPGPKGAAGPCKAGRG-----VPPGAPGAPGAPGAPGAPGAPGAP 762
Db 1061 GPAGPAGAGSPGAPAGDKGAGGQGERGMKGRHGFNDLPGPPGAPGAGHAGQSPGAS 1120
QY 763 GPPGAPGAGRGEGGQPGSPGQGLPGPAGPGEAGKGE---QGVPGDLGAPGSPGPA 819
Db 1121 GPAGRPPGSGSGSPGKDG---NGLPGIPGPPGPRGRTGDVGPAGPAGPAGPAGP 1177
QY 820 GG 821
Db 1178 GG 1179
```

RESULT 11

Q9YIB4

ID Q9YIB4

AC Q9YIB4; PRELIMINARY; PRT; 1450 AA.

DT 01-MAY-1999 (TrEMBLrel. 10, Created)

```
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen.
OS Cynops pyrrhogaster (Japanese common newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX NCBI_TaxID=8330;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Regenerate forelimbs;
RX MEDLINE=99407244; PubMed=10474166;
RX DOI=10.1002/(SICI)1097-0177(199909)216:1<59::AID-DVDY8>3.3.CO;2-2;
RA Asahina K., Ohara M., Yoshizato K.;
RT "Expression of genes of type I and type II collagen in the formation
and development of the blastema of regenerating newt limb.";
RL Dev. Dyn. 216:59-71(1999).
DR EMBL; AB015438; BAA36973.1; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib_collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR01007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; C1g_helix; 4.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS01208; VWF_C; 1.
DR PROSITE; PS0184; VWF_C; 2.
DR PROSITE; PS0184; VWF_C; 2.
KW Collagen.
SQ SEQUENCE 1450 AA; 137563 MW; ABF8A74841B87B7C CRC64;
```

```
Query Match 55.2%; Score 2561.5; DB 2; Length 1450;
Best Local Similarity 52.3%; Pred. No. 9.7e-95;
Matches 517; Conservative 39; Mismatches 261; Indels 171; Gaps 17;

QY 1 GPPGEP-----GPTGLPGPPGGERG-----GP----- 21
Db 192 GEPGEGAAAGALGPRGLPGPPKNGDDGSGKPKRGPPGSGPQAGRLPGTAGLPGMK 251
QY 22 GSRGPPGADGVAGPKGPA-----GERGSPGAPGKSGPGEAGRPGAGLPGAKGLT 72
Db 252 GHRGFGNGLDGAQDNGPAGPKGPPGPNFENGAPQAGPGLPGERGPPGAPGAGAGND 311
QY 73 GSPGSPGDKTGTGPPG-----AGQDGRPPGPPGARGQAGVGMPPGPKGAA 120
Db 312 GSPGAGAPPPGPTGTPGPPGFGAVGAKDAGPQSGRSGEPGQARGEPGA---PGPAGAA 368
QY 121 GEPKACGERGVPPGPPGAVGP---AGKDEAGAGPAGPAGPAGPAGPAGPAGPAGP 177
Db 369 GPSNPGTDCQPGKGTGATGSPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 428
QY 178 GPAGPPGGEAGKPGEQGVPGDLGAPGSPGAPGEPGTLPGPPGGERGSPGSRGPPGAG 237
Db 429 GEPKACGERPVAGVQGPDPGSGEGKRGKSGEPGAPGAPGAPGAPGAPGAPGAPG 488
QY 238 GPKGAPGERSVPAGPKGSPGAGRPGEAGLPGAKGLTGSPGSPGPDGKTGPPGAGQD 297
Db 489 GPKGAPGERSVPAGPKGSGSRPGPGLPGAKGLTGSPGSPGPDGKTGPPGAGQD 548
QY 298 GRPGPPGPPCARGOAGVGMPPGPKGAAGPCKAGRGVPPGPPGAVGPPGKDGAGAG 357
Db 549 GHPGPPGPPCARGOAGVGMPPGPKGAAGPCKAGRGVPPGPPGAVGPPGKDGAGAG 608
QY 358 GPAGPAGRGEGGQPGSPGQGLPGPAGPGEAGKGEQGVPGDLGAPGSPGSGPAGE--- 413
Db 609 GPSGSPGGERGEGQPGSPGQGLPGPAGPGEAGKGEQGVPGDLGAPGSPGSGPAGE 668
```

QY 414 -----PGTGLPGP-----PGRGGSGRGP----- 435
Db GERGGGAGAGGPRGSGPNDGAKGAGAGAGGPRGPGGQMGFGRSGAMPGAK 728
QY 436 -----GADGAGGKPGAGRGSP-----GPAGPKGSGPAGR 468
Db GDRGDAGTKGADGAGKPGDAGRLTGPIGPPGSGAGPKGEGGSGPAGPTGARGSPGR 788
QY 469 GEAGLPGAKGLTGSGS---PGPDGKTGPPAGQDGRPPGPPGAGQAGVMGFPQPK 525
Db GEPGAPGAGICGPPGADGQPGKAGSGDAGPKDAGAPGAGPTGAPGAGNVGAPGPK 848
QY 526 GAAGPGRAGRGVPCGAGVPCAGKDEAGAGQAGPPGAGPAGRGEGGAGSGPQGLP 585
Db GTRGAAGPFGATGPPGAAGRLGPPGSGNVAGPPGPPGKAGKSGRGTGAPGRSGEP 908
QY 586 GPAGPP---GEAGKGEQGVPCDLGAPGSPGAGPPTGLPGPPGERGGSGRFPFAD 642
Db GPAGPPGSGEKSFGSDGAPAGLPGPQIAGQGVVGLFGQGERGFSGLPGPAGEP 968
QY 643 GVAGPKGAGRGSGPAGP-----KGSPPGAGRPGEAGLPGAK-----GLT 684
Db GKQSGPNGRGPPGSGPPGLGPPGCGREGSGPSEGAPRGDGSFGPKGDRGSGPS 1028
QY 685 GSPSGPDGKTGPPGAGQD---CRPPGPPGAGQAGVMGFPGPKGAGSGPCKAGE- 740
Db GPPGAPGAPGAPGPPVGPAGKNGDRGETGAPGAGPAGSGVVRGAPGAPGAGDKGAGEQ 1088
QY 741 -----RGVPPGAGVPCAGKDEAGAGQAGPPGAGPAGRGEGGAGSGPQGLPGPA 792
Db GERGMKGRGNGMGPPGPPGSGSGEAGAPGSGPAGPRGPPGSGSGTKGVNGLPGPI 1148
QY 793 GPPGAGKRGSGQGVPCDLGAPGSPGAG 820
Db GPPGPRGRNGDVGPAGPPGPPGPPG 1176
RESULT 12
Q640B2 PRELIMINARY; PRT; 1449 AA.
AC Q640B2
DT 25-OCT-2004 (TRENBLrel. 28, Created)
DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TRENBLrel. 28, Last annotation update)
DE Hypothetical protein.
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8364;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole body;
RX PubMed=12477932; DOI=10.1073/pnas.242603999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loguallano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RN SEQUENCE FROM N.A.
RC TISSUE=Whole body;
RA Klein S., Gerhard D.S.;
RL Submitted (Sep-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC082718; AAH82718.1; -;
KW Hypothetical protein.
SQ SEQUENCE 1449 AA; 137539 MW; 456639B1687A3B4B CRC64;
Query Match 55.1%; Score 2558.5; DB 2; Length 1449;
Best Local Similarity 53.2%; Pred. No. 1.3e-94;
Matches 508; Conservative 44; Mismatches 268; Indels 135; Gaps 13;
QY 1 GPPGEPGTGLPGPPGRRGGPS-----RFPAGDVAGPKGPA-----GER 42
Db 222 GKPRGERGPPGQAGRLPGTAGLPGMKHGRFNGLDGAKGDSGAPGPKGPGSGEN 281
QY 43 GSPGAPGKSGPAGRGPEAGLPGAKGLTSGSPGPDGKTGPPG-----A 90
Db 282 GAPQGVPRGLPGRRGPPGSPGAGARGNDGAPGAGPNGPTGAGPPGFPFGVGPKGDA 341
QY 91 GQDGRPPGPPGAGQAGVMGFPKGAACEPGKAGRGVPPGAGVGPAGKDGEGAG 150
Db 342 GPQSRGPDGPGGRRGPPGAGQAGAGPSNPNDSGAGAKGATGAGIAGAGFPFGAR 401
QY 151 GPPGAPGAGRGPPGAGSGPFGQLPGPAGPAGGPKGQGVPGDLGAPGSPGAGE 210
Db 402 GAPGQGGAGPPGKNNGPPGACQNGEPGKAGSGPAGVQVPPGSGEGKGRGEP 461
QY 211 GPTGLPGRGPPGRRGPPGAGQAGVMGFPKGAACEPGKAGRGVPPGAGVGPAGKDGEGAG 270
Db 462 GPAGPPGAGRGPPGRRGPPGAGQAGVMGFPKGAACEPGKAGSGEGRRGPPG 521
QY 271 GAKGLTSGSPGPDGKTGPPGAGQAGVMGFPKGAACEPGKAGSGEGRRGPPG 330
Db 522 GAKGLTSGSPGPDGKTGPPGAGQAGVMGFPKGAACEPGKAGSGEGRRGPPG 581
QY 331 GERGVPPGAVGAPGAGKDEAGAGQAGPPGAGPAGRGEGGQAGSGPFGQLPGPAGPGEA 390
Db 582 GERGVPPGAVGAPGAGKDEAGAGQAGPPGAGPAGRGEGGQAGSGPFGQLPGPAGPGEA 641
QY 391 GKPGQGVPCDLGAPGSGPAGE---PGTGLPGPPGRRGPPGAGQAGVMGFPKGA 447
Db 642 GKPGQGVPCDLGAPGSGPAGE---PGTGLPGPPGRRGPPGAGQAGVMGFPKGA 701
QY 448 GERGSPGAPGKSGPGEAGR-----PGEA 471
Db 702 GAPGGQPPGLQGMPPGERGAGGLPGAKGDRDQGVKSGSDGAPGKDGVRGLTGTPGPPGP 761
QY 472 GLPGAK-----GLTSGSPGPPDGKTGPPGAGQDGRPPGPPGARGQ----- 515
Db 762 GAPDKGAGPAGPAGTARGAGPGERGESPPGPPGAGPPGADGQFGAKGEGQDAGAK 821
QY 516 -----AGVMGPPKGAAGEPGKAGRGVPG-----PGAVGAPKDGEGAGAGQPP 561
Db 822 GDAGPPGPPGAGTARGAGPAGALGATGPKGARGAPGATGPPGATGTRVGPAGPAGNSGP 881
QY 562 GPAGPAGE-----RGEQGRAGSGPQGLPGPAGPAGRGEGGQGVPCDLGAPGSGPA 615
Db 882 GPSGAPGKAGKAPRGRTGPPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 941
QY 616 GEPGPTGLPGPPGRRGPPGRRGPPGAGQAGVMGFPKGAACEPGKAGSGPFGAGP---KGSPEGAGR- 671
Db 942 GSRGTGVLPGMRGRRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 1001
QY 672 ---PGEAGLPGAKGLT-----GSPSGPDGKTGPPGAGQDGRPPGPPGPPGPP 714
Db 1002 GAPGSEGAPGRDGAAGPKDRGEGGAPGPPCAPGAPGAPGPPGAPGKSGDRGETGAPGA 1061
QY 715 GARGQAGVMGFPKGAAGEPGKAGE-----RGVPPGPPGAGVGPAGKDGEGAGAGQPP 765
Db 1062 GPAGVAGARGPAGPQGPGRGKGEAGQGERGMKGRGPPGPPGPPGPPGPPGPPGPPGPPG 1121

Qy 766 GPAGPAGERGEQQAPGSPGFQGLPGPAGPBGBAKPKGVGGVPGDI.GAPEGPGFAG 820
||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Dd 1122 GPAGPRGPSSGNPGKDGANGLPFGIPGPPGRGTGDVGPAAGPFPFGPPGPPG 1176
||||| : ||||| : ||||| : ||||| : ||||| : ||||| :

RESULT 13

Q80ZB5 PRELIMINARY; PRT; 1449 AA.

ID Q80ZB5

AC Q80ZB5;

CD AC 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DD 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Collal-prov protein.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

OC xenopodinae; Xenopus.

OX NCBI_TaxID=8355;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Tohiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Viallan D.K., Muzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahay J., Helton E., Kettemann M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skaleka U., Smailus D.E., Schmerch A., Schein J.E.,

RA Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences";

RV Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RX MEDLINE=223441132; PubMed=12454917; DOI=10.1002/dvdy.10174;

RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,

RA Richardson P.i

RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus

RT initiative";

RL Dev. Dyn. 225:384-391(2002).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

CC Klein S., Strausberg R.i

RL Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.

DR ENBL; BC049829; AAH49829.1; -

DR GO; GO:0005581; C:collagen; IEA.

DR GO; GO:0005737; C:cyttoplasm; IEA.

DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.

DR GO; GO:0006817; P:phosphate transport; IEA.

DR InterPro; IPR008161; Clg_helix.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR000885; Fib_collagen_C.

DR InterPro; IPR009041; PMP_SGCI.

DR InterPro; IPR001007; VWFC.

DR Pfam; PF01410; COLFI_1_-

DR Pfam; PF01391; Collagen; 18.

DR Pfam; PF00093; WVC; 1.

DR ProDom; PD000007; Clg_helix; 4.

DR ProDom; PD002078; Fib_collagen_C; 1.

DR SMART; SM00038; COLFI; 1.

RESULT 14
 Q6PCL3 AC Q6PCL3 PRELIMINARY; PRT; 1225 AA.
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Colial protein.
 GN Name=Colial;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udoin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mix FVB/N; TISSUE=Mammary tumor;
 RX Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
 RA EMBL; BC059281; AAHS9281.1; -
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR InterPro; IPR008161; Clq helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR008885; Fib collagen_C.
 DR InterPro; IPR009041; PMP_SGCI.
 DR Pfam; PF01391; Collagen; 13.
 DR Pfam; PF00093; VWC; 1.
 DR ProDom; PD000007; Clq_helix; 2.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWF_C_1; 1.
 DR PROSITE; PS0184; VWF_C_2; 1.
 KW Collagen.
 SQ SEQUENCE 1225 AA; 117860 MW; B6B86CB84457F4D9 CRC64;
 Query Match 53.8%; Score 2494; DB 2; Length 1225;
 Best Local Similarity 56.0%; Pred. No. 4.2e-92;
 Matches 510; Conservative 29; Mismatches 227; Indels 144; Gaps 20;
 QY 12 PGPPCERGG-----PGSRGFFGADGACGKPGAGERSGPGAGKSGSGEAGRPFCEAGLP 66
 DB 73 PNPQRRECCAFCEPEEVVSNSEVDV-GVEGPKDGPQGRGPGVGGPRDGIQFQCLP 131
 QY 67 GAKGLTSGPGSGPDGK-----TGPPGAGQDGRPPGPPGARGQA 108

Db 132 GPPCHPFPFPGLGNGFNFSQMSYGYDEKSAVSVPFPMGSPGPRGLPGPPGARGPQ---- 188
 QY 109 GVMGFPQPKGAAGEPGKAGRGVPGFPGVAGPAGKDGEGAGQAGPP-----GPAGPAGERGEQ 165
 Db 189 ---GFQGFPPGEPGEPGSGPMGPRGPPGPGKNGDDGEGAGKPGRGPRGPPGPGCARGLP 245
 QY 166 GPAGSP-----GFQGLPGPAGPGEAGKPGEGVPGDLGNAPGSPGAGERGPTGLPGPP 219
 Db 246 GTAGLPGMKHRRGFSGLDGAKGADGAPGAPGKPGSPGNGAPGQMGPRGLPGERGRPGPP 305
 QY 220 GERGGPSGRGFPFGADGVAGPKPGAGERSGPGAPGKSGPGEAGRPGEAGLPAGKGLTGPS 279
 Db 306 GT---AGARGNDGAVGAGPGGPT-----GPTGPPGFPAGVAGKAGAGPGGAR----- 350
 QY 280 GSPGPDGKTGPPGPGAGQDGRPPGPPGARGQAGVMGFPFGKGAAGSPGKAGRGVPGPP 339
 Db 351 GSEGPQGVRGEPGPPGAGAAAGPAGNPGADQ-----PGAAGANGAPGIA---GAPGFP 401
 QY 340 GAVGPAKDGEGAGAGQAGPPGAGPAGERGEOGPAGSPGFQGLPGPAGPGEAGKPGEGQVP 399
 Db 402 GARGP-----SGPQSPGPPGPKNSGE-----PGAAGNKGTGAKGEPGATGVQGP 449
 QY 400 GDLGAPGSPGAPGPGTGLPGRGPGSGRGGPGGADGVAGPKGAGERSGPGAPGK 459
 Db 450 GPAGEEKRGAREGPGSGLPGPPGRRGGPGSGRGGPGADGVAGPKGSGERKAGPAPGPK 509
 QY 460 GSPGAGRGPGEAGLPAGKGLTSGSPGPPGKTPGPPGAGQDGRPPGPPGARGQAGVM 519
 Db 510 GSPGAGRGPGEAGLPAGKGLTSGSPGPPGKTPGPPGAGQDGRPPGPPGARGQAGVM 569
 QY 520 GPPGPKAAEPGKAGRGVPGPPGAVGPAKDGEGAGAGQAGPPGAGPAGERGEOGPAGSP 579
 Db 570 GPPGPKTAGEPGKAGRGVPGPPGAVGPAKDGEGAGAGQAGPAGPAGERGEOGPAGSP 629
 QY 580 GFQGLPGPAGPPGAGKPGQGVPGDLGAPGSPGAGE-----PGP----- 620
 Db 630 GFQGLPGPAGPPGAGKPGQGVPGDLGAPGSPGAGE-----PGP----- 689
 QY 621 -----TGLPGRGPGRGGPGSGRGGPGADGVAGPKGAGERSGPGAGPKSGPGE- 668
 Db 690 GAPGNDGAKGDTGAPGAPGSGQAGPGLQMPGERCAAGLPGPKGDRGADGPKGADSGPKD 749
 QY 669 -----ACRPGEAGLPAGKGLTSGSPGPPGKTPGPPGAGQDGRPPGPPGARGQA 720
 Db 750 GARGLTGPIGPPGPGAGAGDKGEGAGSGPPGPTGARGAPGDRGAGPPGAGPAGPCAP 809
 QY 721 GVMGFPKGAAGEPGKAGRGVPGPPGAVGPAKDGEGAGAGQAGPPGAGPAGERGE----- 776
 Db 810 GAPGAPGVGPAKNGDRGETGPAGPAGPIGPAARGPAGPQGRGDKGETGEGQDRIK 869
 QY 777 -----QGPAGSPGPGQLPGPA-----GPPGEGAGKPGE---QGVPGDLGAPGP- 815
 Db 870 GHRGFSGLQPPGSPGSGGSGAGSPAGPRGPPGSGAGSPGKDLNGLPGIPGPPGPR 929
 QY 816 -----SGPAG 820
 Db 930 GRTGDSGPAG 939
 RESULT 15
 Q14046 PRELIMINARY; PRT; 1160 AA.
 ID Q14046
 AC Q14046;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE COL2A1 protein precursor (fragment).
 GN Name=COL2A1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:15:59 ; Search time 88.8873 Seconds
(without alignments)
2367.017 Million cell updates/sec

Title: US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSECPGVGEGPFGPPGAGA.....PGSGDAGPPGPPGACKEG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq.l6Dec04.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3070	100.0	544	ADM48393	Adm48393 Recombina
2	2966	96.6	1057	AAy84541	AAy84541 Amino aci
3	2966	96.6	1057	AAy84544	AAy84544 A human c
4	2966	96.6	1058	AAy84403	AAy84403 Amino aci
5	2966	96.6	1107	AAr89472	AAr89472 Collagen/
6	2966	96.6	1107	AAy84540	AAy84540 Amino aci
7	2966	96.6	1161	AAE87050	AAE87050 Human pan
8	2966	96.6	1169	AAr89469	AAr89469 Collagen/
9	2966	96.6	1169	AAy84537	AAy84537 Amino aci
10	2966	96.6	1171	AAr89470	AAr89470 Collagen/
11	2966	96.6	1171	AAy84538	AAy84538 A chimeri
12	2966	96.6	1211	AAE87057	AAE87057 Human pan
13	2966	96.6	1226	AAE87062	AAE87062 Human pan
14	2966	96.6	1388	AAr89471	AAr89471 Collagen/
15	2966	96.6	1411	AAy56800	AAy56800 Human pre
16	2966	96.6	1461	ABg93947	ABg93947 Human pol
17	2966	96.6	1464	AAw68485	AAw68485 Human rec
18	2966	96.6	1464	AAE82454	AAE82454 Human pro
19	2966	96.6	1464	AAU14136	AAU14136 Human nov
20	2966	96.6	1464	AAy50764	AAy50764 Human tum
21	2966	96.6	1464	ABP68610	ABP68610 Human pan
22	2966	96.6	1464	ABU54471	ABU54471 Human tum
23	2966	96.6	1464	ABR47417	ABR47417 Breast ca
24	2966	96.6	1464	ABR92064	ABR92064 Human cer
25	2966	96.6	1464	ADD14142	ADD14142 Human src

26	2966	96.6	1464	7	ADD45059	ADD45059 Human Pro
27	2966	96.6	1464	7	ADD45055	ADD45055 Human Pro
28	2966	96.6	1464	7	ADD45051	ADD45051 Human Pro
29	2966	96.6	1464	7	ADP87048	ADP87048 Human alp
30	2966	96.6	1464	7	ADP65246	ADP65246 Human alp
31	2966	96.6	1464	8	ADQ19470	ADQ19470 Human sof
32	2966	96.6	1464	8	ADQ29653	ADQ29653 Human col
33	2966	96.6	1464	8	ADP16800	ADP16800 Human col
34	2966	96.6	1464	8	ADR16425	ADR16425 Human col
35	2966	96.6	1464	8	ADR99144	ADR99144 Collagen,
36	2958	96.4	1388	3	AAy84539	AAy84539 Amino aci
37	2936	95.6	1463	4	AAE02532	AAE02532 Bovine al
38	2920	95.1	1536	7	ADP87051	ADP87051 Human pan
39	2872	93.6	1453	7	ADD45053	ADD45053 Rat Prote
40	2872	93.6	1453	7	ADD45057	ADD45057 Rat Prote
41	2872	93.6	1453	7	ADD48341	ADD48341 Rat Prote
42	2872	93.6	1453	7	ADD45049	ADD45049 Rat Prote
43	2872	93.6	1453	7	ADD48337	ADD48337 Rat Prote
44	2872	93.6	1453	7	ADD48345	ADD48345 Rat Prote
45	2857	93.1	822	2	AAy06240	AAy06240 Mouse rec

ALIGNMENTS

RESULT 1
ADM48393
ID ADM48393 standard; protein; 544 AA.
XX
AC ADM48393;
XX
AC
DT 03-JUN-2004 (first entry)
XX
DE Recombinant gelatin-like polypeptide Hu-deam.
XX
KW Plasma substitute; Gelatin-like protein; plasma expander.
XX
OS Synthetic.
XX
FN EPI398324-A1.
XX
PD 17-MAR-2004.
XX
PF 11-SEP-2002; 2002EP-00078745.
XX
PR 11-SEP-2002; 2002EP-00078745.
XX
PA (FUJF) FUJI PHOTO FILM BV.
XX
PI Bouwstra JB, Toda Y;
XX
DR WPI, 2004-229415/22.
XX
PT Composition useful as substitute for plasma, comprises solution of saline
and recombinant gelatin-like protein having colloid osmotic function.
XX
Example 1; SEQ ID NO 4; 31pp; English.

The present sequence is the protein sequence of recombinant gelatin-like protein Hu-deam. A claimed composition suitable as a substitute for plasma comprises a solution of saline and a protein having a colloid osmotic function. The protein is a recombinant gelatin-like protein with a molecular weight of at least 10 kDa and at most 50 kDa and an isoelectric point of less than 8. It is especially Hu-1 ADM48393 or Hu-deam. The recombinant gelatin-like protein, or a dimer, trimer or tetramer of the protein, is useful as a plasma expander that has a lower clearance rate from blood circulation, provides better and predictable regulation of clearance rate and which is less susceptible to proteolytic degradation than presently used gelatin derivatives. Recombinant gelatin-like proteins that are in essence free of hydroxyproline do not give rise to an immunological reaction with blood samples containing Ige antibodies. The gelatin-like proteins can be covalently attached to pharmaceutical active compounds. After administration, the coupled

Db 261 GPGGPPGKNSGEGPCAGSGKDTCAKGEPPGVGQPPGAGBEGKRGARGEPGPTGLP 320
 Qy 121 GPPGEGGSGRGGPCADGAVGPKGAGERSGPGAGPKGSPGEGAGRPGEAGLPCAKGLT 180
 Db 321 GPPGEGGSGRGGPCADGAVGPKGAGERSGPGAGPKGSPGEGAGRPGEAGLPCAKGLT 380
 Qy 181 GSPSGPDGKTGTPGAGDGRPCPPGARGGAGVGMGPKGAAGEPKAGRGV 240
 Db 381 GSPSGPDGKTGTPGAGDGRPCPPGARGGAGVGMGPKGAAGEPKAGRGV 440
 Qy 241 GPPGAVGAGKDGAGAGGPPGPPGAGPAGERGEGGPGAGSPGEGLPAGPPGEGAGKPGEE 300
 Db 441 GPPGAVGAGKDGAGAGGPPGPPGAGPAGERGEGGPGAGSPGEGLPAGPPGEGAGKPGEE 500
 Qy 301 GVPDGLGAPGSGARGGPPGGERGVGPPGAGPPGADGAPGDDGAKGDAGAPGSGE 360
 Db 501 GVPDGLGAPGSGARGGPPGGERGVGPPGAGPPGADGAPGDDGAKGDAGAPGSGE 560
 Qy 361 GAPLEGMPGEGAGLPGPKDGDAGPKGADGSPGKDGVRGLTGTPTGPPGAGAGDK 420
 Db 561 GAPLEGMPGEGAGLPGPKDGDAGPKGADGSPGKDGVRGLTGTPTGPPGAGAGDK 620
 Qy 421 GESGSGPAGTGAAGAPGDRGEGPPGPPGAGPPGADGAPGDDGAKGDAGAGDK 480
 Db 621 GESGSGPAGTGAAGAPGDRGEGPPGPPGAGPPGADGAPGDDGAKGDAGAGDK 680
 Qy 481 GPAGPAGPPGTPGVGAPGAKGARGSGAGPPGATGPPGAAGRVGPPGSGNAGPPGPPGPA 540
 Db 681 GPAGPAGPPGTPGVGAPGAKGARGSGAGPPGATGPPGAAGRVGPPGSGNAGPPGPPGPA 740
 Qy 541 GKEG 544
 Db 741 GKEG 744

RESULT 6
 RAY84540

ID AAY84540 standard; protein; 1107 AA.
 AC AAY84540;
 XX
 XX 25-JUL-2000 (first entry)
 DT Amino acid sequence of a chimeric collagen 1 (alpha1)/decorin protein.
 DE Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibronectin; fibronectin; post translational hydroxylation;
 KW decorin; chimera.
 XX Homo sapiens.
 OS Unidentified.
 OS Chimeric.
 XX Key Location/Qualifiers
 FT Misc-difference 858
 FT /note= "Gly encoded by GCT"
 XX
 XX EP992586-A2.
 XX
 XX 12-APR-2000.
 PD
 XX 07-OCT-1999; 99EP-00119184.
 XX
 XX 09-OCT-1998; 98US-00169768.
 PR (USSU) US SURGICAL CORP.
 XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
 PI WPI: 2000-259138/23.
 DR N-PSDB; AAI12500.
 XX

PT Production of extracellular matrix proteins containing 4-trans-
 FT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.
 XX Claim 24; Fig 18; 260pp; English.
 XX The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibronectin and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a chimeric collagen 1 (alpha1)/decorin
 CC protein, which may be produced using the method of the invention
 XX
 SQ Sequence 1107 AA;

Query Match 96.6%; Score 2966; DB 3; Length 1107;
 Best Local Similarity 95.6%; Pred. No. 3e-169;
 Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 GSEGEVGRGEPGPPGAGAGPAGDPCADGEPGAKGADGAPGAGPFGGARGSPGPE 60
 Db 201 GSEGPQVRGEPGPPGAGAGPAGDPCADGEPGAKGADGAPGAGPFGGARGSPGPE 260
 Qy 61 GPGGPPGKDSGEPGAPGSKGDTGAKGEPGVGVEGPPGAGGEGKRGARGEPGPTGLP 120
 Db 261 GPGGPPGKDSGEPGAPGSKGDTGAKGEPGVGVEGPPGAGGEGKRGARGEPGPTGLP 320
 Qy 121 GPPGEGGSGRGGPCADGAVGPKGAGERSGPGAGPKGSPGEGAGRPGEAGLPCAKGLT 180
 Db 321 GPPGEGGSGRGGPCADGAVGPKGAGERSGPGAGPKGSPGEGAGRPGEAGLPCAKGLT 380
 Qy 181 GSPSGPDGKTGTPGAGDGRPCPPGARGGAGVGMGPKGAAGEPKAGRGV 240
 Db 381 GSPSGPDGKTGTPGAGDGRPCPPGARGGAGVGMGPKGAAGEPKAGRGV 440
 Qy 241 GPPGAVGAGKDGAGAGGPPGPPGAGPAGERGEGGPGAGSPGEGLPAGPPGEGAGKPGEE 300
 Db 441 GPPGAVGAGKDGAGAGGPPGPPGAGPAGERGEGGPGAGSPGEGLPAGPPGEGAGKPGEE 500
 Qy 301 GVPDGLGAPGSGARGGPPGGERGVGPPGAGPPGADGAPGDDGAKGDAGAPGSGE 360
 Db 501 GVPDGLGAPGSGARGGPPGGERGVGPPGAGPPGADGAPGDDGAKGDAGAPGSGE 560
 Qy 361 GAPLEGMPGEGAGLPGPKDGDAGPKGADGSPGKDGVRGLTGTPTGPPGAGAGDK 420
 Db 561 GAPLEGMPGEGAGLPGPKDGDAGPKGADGSPGKDGVRGLTGTPTGPPGAGAGDK 620
 Qy 421 GESGSGPAGTGAAGAPGDRGEGPPGPPGAGPPGADGAPGDDGAKGDAGAGDK 480
 Db 621 GESGSGPAGTGAAGAPGDRGEGPPGPPGAGPPGADGAPGDDGAKGDAGAGDK 680
 Qy 481 GPAGPAGPPGTPGVGAPGAKGARGSGAGPPGATGPPGAAGRVGPPGSGNAGPPGPPGPA 540
 Db 681 GPAGPAGPPGTPGVGAPGAKGARGSGAGPPGATGPPGAAGRVGPPGSGNAGPPGPPGPA 740
 Qy 541 GKEG 544
 Db 741 GKEG 744

CC A fusion protein (AAR89469) comprises the alpha-helical region of human
CC collagen I(a) linked to the human mature bone morphogenic protein 2B
CC (BMP2B). It can be expressed in *Escherichia coli* transformants carrying a
CC vector incorporating a chimeric gene (AAT16515) coding for the fusion.
CC The BMP moiety induces osteogenesis, while the collagen moiety provides
CC an integral substratum or scaffolding for the BMP and cells involved in
CC reconstruction and growth. The fusion protein provides sustained release
CC and delivery of BMP to a target tissue
XX
SO Sequence 1169 AA:

Query Match	96.6%; Score 2966; DB 2; Length 1169;
Best local Similarity	95.6%; Pred. No. 3.1e-163;
Matches	520; Conservative 21; Mismatches 3; Indels 0; Gaps 0
Qy	1 GSEGEVGRGPPGPPGAGAAPGADPCADCEPGAKGADGAPGIAGAPGFCGARGSPGE 60
Db	201 GSEGEVGRGPPGPPGAGACGAGNPGADQCPGAKGANGAPGIACAPGFCGARGSPGP 260
Qy	61 GPGGPPGPKDGSBPGAPSGKDGTCAGKEGPGVGVGPPGAGEGKPCARGEPGPTGLP 120
Db	261 GPGGPPGPKGNSGEFAPGSGKDGTCAGKEGPGVGVGPPGAGEGKRGARGEPGPTGLP 320
Qy	121 GPPERGGPSRGPPGADGVAGPKPAGERSGPPAGPKGSPGEAGRPGEAGLPCAAGLT 180
Db	321 GPPERGGPSRGPPGADGVAGPKPAGERSGPPAGPKGSPGEAGRPGEAGLPCAAGLT 380
Qy	181 GSPGSPGDGKTGPPGAGEGRCRPGPPGARGAEGVMGPPGPKAAAGEPCKAGERGVP 240
Db	381 GSPGSPGDGKTGPPGAGQDGRPGPPGPPGARGAQVMGPPGPKAAAGEPCKAGERGVP 440
Qy	241 GPPGAVGPKDGBAGABGPPGPPGAPGAGEEGPAGSPGPEGLPGPAGPGCEAKGPGE 300
Db	441 GPPGAVGPKDGBAGAQGPPGPPGAPGAGEGQGPAGSPGQGLPGPAGPGCEAKGPGE 500
Qy	301 GVPDLAGAPGSGARGEPGFPGERGVEGPPGAPGPPGADGAPDDGAKGDGACAPGSGE 360
Db	501 GVPDLAGAPSGARGERGFPGERGVQPPGPPGPRGANGAPNDGAKGDGACAPGPGSQ 560
Qy	361 GAPLEGMPGSRGAAGLPKPKDGRDAGPKGADGSPKDGVRGLTGPIGPPGAPGAGDK 420
Db	561 GAPGLQMPGSRGAAGLPKPKDGRDAGPKGADGSPKDGVRGLTGPIGPPGAPGAGDK 620
Qy	421 GESGSPGAPGTGARGAPGDRGEPGPPGAPGPPGADGPGAGEPCDAGAKGDAGPP 480
Db	621 GESGSPGAPGTGARGAPGDRGEPGPPGAPGPPGADGPGAGEPCDAGAKGDAGPP 680
Qy	481 GPAGPAGPPGPIGDVGPAGKARGSAGPPGATGPPGAAGRVGPPGPGSDAGPPGPPGPA 540
Db	681 GPAGPAGPPGPIGNVGAPKARGSAGPPGATGPPGAAGRVGPPGPGSGNAGPPGPPGPA 740
Qy	541 GKEG 544
Db	741 GKEG 744

RESULT 9	
AAAY84537	
ID	AAAY84537 standard; protein; 1169 AA.
XX	
XX	AAAY84537;
XX	
XX	
DT	25-JUL-2000 (first entry)
XX	
DE	Amino acid sequence of a chimeric collagen 1 (alpha1)/BMP-2B protein.
XX	
XX	Extracellular matrix protein; self aggregation; hydroxylated proline;
KW	trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW	collagen; fibrinogen; fibronectin; post translational hydroxylation;
XX	ss. bone morphogenic protein; BMP-2B; chimera.
XX	
OS	Homo sapiens.
OS	Unidentified.

OS	Chimeric.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 677	
FT		/note= "Ala encoded by G"
FT	Misc-difference 887	
FT		/note= "unspecified amino acid encoded by CT"
FT	Misc-difference 890	
FT		/note= "unspecified amino acid encoded by CT"
XX		
PN	EP92586-A2.	
XX		
XX	12-APR-2000.	
PD		
XX		
XX	07-OCT-1999;	99EP-00119184.
XX		
PR	09-OCT-1998;	98US-00169768.
XX		
XX	(USSU)	US SURGICAL CORP.
PA		
XX		
PI	Gruskin EA, Buechter DD, Zhang G, Connolly K;	
XX		
DR	WPI; 2000-259138/23.	
DR	N-PSDB; AAA12497.	
XX		
PT	Production of extracellular matrix proteins containing 4-trans-	
PT	hydroxyproline results in native self aggregating proteins, useful on	
PT	medical implants.	
XX		
PS	Claim 22; Fig 13; 260pp; English.	

The specification describes a method for producing an extracellular matrix protein or its fragment. The extracellular matrix protein is capable of self aggregating in a cell which does not ordinarily hydroxylate prolines. The method comprises optimising a nucleic acid sequence for expression in the cell by substitution of codons preferred by that cell for naturally occurring codons not preferred by the cell; incorporating the nucleic acid sequence into the cell; and contacting the cell with a hypertonic growth medium containing at least one amino acid, selected from the group consisting of trans-4-hydroxyproline and 3-hydroxyproline to allow at least one of the amino acids to be assimilated into the cell and incorporated into the extracellular matrix protein. The method may be used to make host cells assimilate and incorporate trans-4-hydroxyproline into proteins. This is especially useful in the recombinant production of proteins such as collagen, fibronogen and fibronectin whose ability to self aggregate and produce functional proteins depends on the post translational hydroxylation of proline. The method is also useful in studying the structure and function of polypeptides which do not normally contain trans-4 (alpha)hydroxyproline. The present sequence represents a chimeric collagen 1 (alpha1)I/bone morphogenic protein-2B (bmp-2b) protein, which may be produced using the method of the invention

Seq	Sequence	1169	AA;
	Query Match	96.6%	Score 2966; DB 3; Length 1169;
	Best Local Similarity	95.6%	Pred. No. 3.1e-169;
	Matches	520;	Conservative 21; Mismatches 3; Indels 0; Gaps 0
Qy	1	GSEGPQVGRCEP	PPGPPAGAAAGDPCGADGEPGCAKGADGAPGIAGAPFFCARGSPSPGE 60
Db	201	GSEGPQVGRCEP	PPGPPAGAAAGDPCGADGEPGCAKGADGAPGIAGAPFFCARGSPSPGE 260
Qy	61	GPQGPQPCGDS	GEPPGAPGSKGDTGAKGEPGVGEGPPGPAGGEGKPGARGEPGPTGLP 120
Db	261	GPQGPQPCGDS	GEPPGAPGSKGDTGAKGEPGVGEGPPGPAGGEGKPGARGEPGPTGLP 320
Qy	121	GPPEERGSGSR	FFPCGADGVAGPKPGAGERGSGPPAGPKGSPGAGRPCEAGLPCAKGLT 180
Db	321	GPPEERGSGSR	FFPCGADGVAGPKPGAGERGSGPPAGPKGSPGAGRPCEAGLPCAKGLT 380
Qy	181	GSPGSGPDGKT	GPFGPPAGBDRGPPGPPGARGGAGVMGPPGPKAAGEFPKAGERGVP 240

Db 381 CSPSPGPDGKTPPPGPAQDGRPPGPPGARGQAGVMGPPGPKGAGGPGKAGRGVP 440
 QY 241 GPPGAVGAGKDGAGAGAGPPGPPGAGGEEGPPGAGSGPGEGLPGPAGPPGAGKPGEE 300
 Db 441 GPPGAVGAGKDGAGAGAGPPGPPGAGGEEGPPGAGSGPGEGLPGPAGPPGAGKPGEE 500
 QY 301 GVPDGLGAPSGARGGPPGPPGAGGEEGPPGAGSGPGEGLPGPAGPPGAGKPGEE 360
 Db 501 GVPDGLGAPSGARGGPPGPPGAGGEEGPPGAGSGPGEGLPGPAGPPGAGKPGEE 560
 QY 361 GAPLEGMPGRRGAGLPPGKDRGDAGPKGADSGPKGVRLTGTGIPGPPGAGAPGDK 420
 Db 561 GAPLGQMPGRRGAGLPPGKDRGDAGPKGADSGPKGVRLTGTGIPGPPGAGAPGDK 620
 QY 421 GESGSPGAGTARGAPGPPGPPGAGGEEGPPGAGSGPGEGLPGPAGPPGAGKPGEE 480
 Db 621 GESGSPGAGTARGAPGPPGPPGAGGEEGPPGAGSGPGEGLPGPAGPPGAGKPGEE 680
 QY 481 GPAGPAGPPGPIGVDGAPGAGKAGSGAPGATGPPGAAAGRVGPPGSGDAGPPGPPGA 540
 Db 681 GPAGPAGPPGPIGVDGAPGAGKAGSGAPGATGPPGAAAGRVGPPGSGDAGPPGPPGA 740
 QY 541 GKEG 544
 Db 741 GKEG 744

RESULT 10
 AAR89470
 ID AAR89470 standard; protein; 1171 AA.
 AC AAR89470;
 DT 01-OCT-1996 (first entry)
 DE Collagen/TGF-beta-1 fusion protein.
 KW Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;
 KW bone formation; tissue repair; fusion protein.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Domain 1..1057
 FT /label= Collagen-IA
 FT /note= "collagen IA alpha-helical domain"
 FT Misc-difference 887
 FT /note= "unidentified amino acid"
 FT Misc-difference 890
 FT /note= "unidentified amino acid"
 FT Peptide 1058..1059
 FT /label= linker_peptide
 FT Domain 1060..1171
 FT /label= TGF-beta-1
 FT /note= "human mature TGF-beta-1"
 XX CA215147-A.
 XX
 XX
 XX
 XX 11-DEC-1995.
 XX
 XX 12-JUN-1995; 95CA-02151547.
 XX
 XX 10-JUN-1994; 94US-00259263.
 XX
 XX (USSU) US SURGICAL CORP.
 XX
 XX Gruskin EA, Espino P;
 XX
 XX WPI; 1996-140144/15.
 XX N-PSDB; AAT16516.
 XX
 XX Chimaeric DNA encoding protein contg. extracellular matrix protein domain
 PT - and cellular regulatory factor domain, partic. useful as osteogenic

PT agents, also related vectors, transformed cells and chimaeric proteins.
 XX Disclosure; Fig 6; 59pp; English.
 CC A fusion protein (AAR89470) comprises the alpha-helical region of human
 CC collagen I(a) linked to the human mature transforming growth factor beta-
 CC 1 (TGF-beta-1). It can be expressed in Escherichia coli transformants
 CC carrying a vector incorporating a chimeric gene (AAT16516) coding for the
 CC fusion. The TGF-beta- moiety increases efficacy of the body's normal soft
 CC tissue repair response and also induces osteogenesis. The collagen moiety
 CC provides an integral substratum or scaffolding for the TGF and cells
 CC involved in reconstruction and growth. The fusion protein provides
 CC sustained release and delivery of TGF-beta-1 to a target tissue
 XX
 SQ Sequence 1171 AA;
 Query Match 96.6%; Score 2966; DB 2; Length 1171;
 Best Local Similarity 95.6%; Pred. No. 3.1e-169;
 Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GSEGGVGRGPPGPPGAGAGPAGDPGADGEPGAKGADGAPGIAGAPFPGAGPSGP 60
 Db 201 GSEGGVGRGPPGPPGAGAGPAGDPGADGEPGAKGADGAPGIAGAPFPGAGPSGP 260
 QY 61 GPGGPPGKDSGPPGAPGSKGDTGAKGPPGVGVEGPPGAGGEGKPGARGPCPTGLP 120
 Db 261 GPGGPPGKDSGPPGAPGSKGDTGAKGPPGVGVEGPPGAGGEGKPGARGPCPTGLP 320
 QY 121 GPPHRRGPPGRRPPGADGVAGPKGPPGAGERSGPPGPKGSPGAGRPGEAGLFGAKGLT 180
 Db 321 GPPHRRGPPGRRPPGADGVAGPKGPPGAGERSGPPGPKGSPGAGRPGEAGLFGAKGLT 380
 QY 181 GSPGSPGPDGKTGPPGAGDGRGPPGPPGAGGEGKPGARGPCPTGLP 240
 Db 381 GSPGSPGPDGKTGPPGAGDGRGPPGPPGAGGEGKPGARGPCPTGLP 440
 QY 241 GPPGAVGPPGKDSGPPGAGGPPGPPGAGGEGKPGARGPCPTGLP 300
 Db 441 GPPGAVGPPGKDSGPPGAGGPPGPPGAGGEGKPGARGPCPTGLP 500
 QY 301 GVPDGLGAPSGARGGPPGPPGAGGEEGPPGAGSGPGEGLPGPAGPPGAGKPGEE 360
 Db 501 GVPDGLGAPSGARGGPPGPPGAGGEEGPPGAGSGPGEGLPGPAGPPGAGKPGEE 560
 QY 361 GAPLEGMPGRRGAGLPPGKDRGDAGPKGADSGPKGVRLTGTGIPGPPGAGAPGDK 420
 Db 561 GAPLGQMPGRRGAGLPPGKDRGDAGPKGADSGPKGVRLTGTGIPGPPGAGAPGDK 620
 QY 421 GESGSPGAGTARGAPGPPGPPGAGGEEGPPGAGSGPGEGLPGPAGPPGAGKPGEE 480
 Db 621 GESGSPGAGTARGAPGPPGPPGAGGEEGPPGAGSGPGEGLPGPAGPPGAGKPGEE 680
 QY 481 GPAGPAGPPGPIGVDGAPGAGKAGSGAPGATGPPGAAAGRVGPPGSGDAGPPGPPGA 540
 Db 681 GPAGPAGPPGPIGVDGAPGAGKAGSGAPGATGPPGAAAGRVGPPGSGDAGPPGPPGA 740
 QY 541 GKEG 544
 Db 741 GKEG 744

RESULT 11
 AAY84538
 ID AAY84538 standard; protein; 1171 AA.
 XX
 XX AAY84538;
 AC
 XX
 XX 25-JUL-2000 (first entry)
 DT
 XX A chimeric collagen 1 (alpha1)/TGF-beta1 protein.
 DE
 XX Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;

KW	collagen; fibronectin; fibronectin; post translational hydroxylation;
KX	ss. transforming growth factor-beta1; TGF-beta1; chimera.
OS	Homo sapiens.
OS	Unidentified.
OS	Chimeric.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 858
FT	/note= "Gly encoded by GCT"
XX	
PN	EP992586-A2.
XX	
PD	12-APR-2000.
XX	
PF	07-OCT-1999; 99EP-00119184.
XX	
PR	09-OCT-1998; 98US-00169768.
XX	
PA	(USSU) US SURGICAL CORP.
XX	
PI	Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX	
DR	WPI; 2000-259138/23.
DR	N-PSDB; AAA12498.
XX	
PT	Production of extracellular matrix proteins containing 4-trans-
PT	hydroxyproline results in native self aggregating proteins, useful on
PT	medical implants.
XX	
PS	Claim 23; Fig 15; 260pp; English.
XX	
CC	The specification describes a method for producing an extracellular
CC	matrix protein or its fragment. The extracellular matrix protein is
CC	capable of self aggregating in a cell which does not ordinarily
CC	hydroxylated prolines. The method comprises optimising a nucleic acid
CC	sequence for expression in the cell by substitution of codons preferred
CC	by that cell for naturally occurring codons not preferred by the cell;
CC	incorporating the nucleic acid sequence into the cell; and contacting the
CC	cell with a hypertonic growth medium containing at least one amino acid,
CC	selected from the group consisting of trans-4-hydroxyproline and 3-
CC	hydroxyproline to allow at least one of the amino acids to be assimilated
CC	into the cell and incorporated into the extracellular matrix protein. The
CC	method may be used to make host cells assimilate and incorporate trans-4-
CC	hydroxyproline into proteins. This is especially useful in the
CC	recombinant production of proteins such as collagen, fibronectin and
CC	fibronectin whose ability to self aggregate and produce functional
CC	proteins depends on the post translational hydroxylation of proline. The
CC	method is also useful in studying the structure and function of
CC	polypeptides which do not normally contain trans-4-hydroxyproline. The
CC	present sequence represents chimeric collagen 1 (alpha1)/transforming
CC	growth factor-beta1 (TGF-beta1) protein, which may be produced using the
CC	method of the invention
XX	
SQ	Sequence 1171 AA;
	Query Match 96.6%; Score 2966; DB 3; Length 1171;
	Best Local Similarity 95.6%; Pred No. 3,1e-169;
	Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0
QY	1 GSEGEVGRGEPGPPGAGAGPAGDPGADGPGAGKADGAFGIAGAPFGPAGFGSPGE 60
DB	: : : : : : : : : :
DB	201 GSEGFQGVGRGEPGPPGAGAGPAGDPGADGPGAGKADGAFGIAGAPFGPAGFGSPGE 260
QY	61 GPGGPPGPKGDSGEFGAPGSKGDTGAKGPPGVGVGPPGPPGABEGKPGARPEGPTGLP 120
DB	: : : : : : : : : :
DB	261 GPGGPPGPKGNSGEFGAPGSKGDTGAKGPPGVGVGPPGPPGABEGKPGARPEGPTGLP 320
QY	121 GPPGERGGPGSRGFFPGADGVAGPKGAPGRGSPGAPGKSPGEGAGRPGCEAGLPGAKGILT 180
DB	: : : : : : : : : :
DB	321 GPPGERGGPGSRGFFPGADGVAGPKGAPGRGSPGAPGKSPGEGAGRPGCEAGLPGAKGILT 380
QY	181 GSPGSGPGDGKTTGPPCPAGGEDGRGPPGPPGARGAGVGMFPFGKAGGEPGKAGRGVPP 240

Db	381	GSFGSGPDGKTGPPGPAQDGRFGPPPPARQAQVWFPFGPKGAAGFPKAGRGVP	440			
QY	241	GPAGVAGVAGKDGAGAGPPGPAAGERGERGEPAGSPGEGLPGPAGPPGAGKPGEE	300			
Db	441	GPAGVAGVAGKDGAGAGPPGPAAGERGERGEPAGSPGEGLPGPAGPPGAGKPGEE	500			
QY	301	GVPGDLGAPGSGARGEFGFGERGVGPPGAPGADGAPDDGAKGDAGAPGAPGSE	360			
Db	501	GVPGDLGAPGSGARGEFGFGERGVGPPGAPGADGAPDDGAKGDAGAPGAPGSE	560			
QY	361	GAPGLEHWPGERGAAGLPCPKDGDAGPKGADGSPKGVGRLTGPIGPPGAPGAPGDK	420			
Db	561	GAPGLEHWPGERGAAGLPCPKDGDAGPKGADGSPKGVGRLTGPIGPPGAPGAPGDK	620			
QY	421	GSFGSGPAGPTGARGAPDGRGEPGPPGAPGADGEPKAGKPPGPDAGAKGDAGPP	480			
Db	621	GSFGSGPAGPTGARGAPDGRGEPGPPGAPGADGEPKAGKPPGPDAGAKGDAGPP	680			
QY	481	GPAGPAGPPGPTGVDGAPGAKGARGSGPPGATGPPGAAGRVPGPSGDAGPPGPPGA	540			
Db	681	GPAGPAGPPGPTGVDGAPGAKGARGSGPPGATGPPGAAGRVPGPSGDAGPPGPPGA	740			
QY	541	GKEG 544				
Db	741	GKEG 744				
RESULT 12						
ADEB87057						
ID	ADEB87057 standard; protein; 1211 AA.					
AC	ADEB87057;					
DT	25-JAN-2004 (first entry)					
DE	Human pancreatic cell protein sequence SeqID517.					
KW	neoplastic pancreatic cell; pancreatic cell; pancreatic cancer;					
KW	cancer death; cytostatic; vaccine; gene therapy;					
KW	non-cancerous pancreas disease; human.					
OS	Homo sapiens.					
PN	WO2003060145-A2.					
PD	24-JUL-2003.					
PF	19-DEC-2002; 2002WO-US040655.					
PR	21-DEC-2001; 2001US-0342768P.					
PA	(DIAD-) DIADEXUS INC.					
PI	Sun Y, Liu C;					
DR	WPI; 2003-587286/55.					
XX	N-PSDB; ADEB87397.					
New pancreatic specific nucleic acid molecule or protein for diagnosing, staging, imaging, monitoring, preventing or treating pancreatic cancer or non-cancerous disease states of the pancreas.						
Claim 12; SEQ ID NO 517; 635pp; English.						
This invention relates to novel nucleic acids and proteins present in normal and neoplastic pancreatic cells. Pancreatic cancer is a common cause of cancer death worldwide, therefore accurate methods of diagnosis and treatment are required. Compounds which modulate the proteins of the invention may have cytostatic activity and the protein and DNA sequences of the invention may be useful for the development of a vaccine or in gene therapy. The composition and methods are useful in diagnosing, staging, imaging, monitoring, preventing or treating pancreatic cancer.						


```

AAR89471
ID AAR89471 standard; protein; 1388 AA.
XX
AC AAR89471;
XX
DT 01-OCT-1996 (first entry)
XX
DE Collagen/decorin fusion protein.
XX
KW Transforming growth factor; TGF-beta-1; collagen IA; osteogenesis;
KW bone formation; tissue repair; fusion protein.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Domain 1..1057
FT /label= Collagen-IA
FT /note= "collagen IA alpha-helical domain"
FT Misc-difference 887
FT /note= "unidentified amino acid"
FT Misc-difference 890
FT /note= "unidentified amino acid"
FT Peptide 1058..1059
FT /label= Linker_peptide
FT Domain 1060..1388
FT /label= Decorin
XX
PN CA2151547-A.
XX
XX 11-DEC-1995.
XX
XX 12-JUN-1995; 95CA-02151547.
XX
XX 10-JUN-1994; 94US-00259263.
XX (USSU ) US SURGICAL CORP.
XX
XX Gruskin EA, Espino P;
XX WPI; 1996-140144/15.
XX N-PSDB; AAT16517.
XX
XX Chimaeric DNA encoding protein contg. extracellular matrix protein domain
FT - and cellular regulatory factor domain, partic. useful as osteogenic
FT agents, also related vectors, transformed cells and chimaeric proteins.
XX
PS Disclosure; Fig 7; 59pp; English.
XX
XX A fusion protein (AAR89471) comprises the alpha-helical region of human
XX collagen I(a) linked to human dermatan sulphate proteoglycan (decorin).
XX It can be expressed in Escherichia coli transformants carrying a vector
XX incorporating a chimeric gene (AAT16517) coding for the fusion. The
XX decorin binds to type I collagen and thus affects Elbril formation. It
XX inhibits the cell attachment-promoting activity of collagen and
XX fibrinogen by binding to such molecules near their cell binding sites.
XX The collagen moiety provides an integral substratum or scaffolding for
XX the decorin. The fusion protein acts to reduce scarring of healing tissue
XX
SQ Sequence 1388 AA;

Query Match          96.6%; Score 2966; DB 2; Length 1388;
Best Local Similarity 95.6%; Pred. No. 3.6e-169;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGEVGRGPPGPPAGAGPAGDPCADGEPGAKGADGAPGTAGAPGPPGARGPSGPE 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 201 GSEGPQGVGRGPPGPPAGAGPAGNPGADGQPCAKGANGAPGTAGAPGPPGARGPSGQ 260
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 GPGGPPGPKSGRPGAPSGKGDGTAKGPGPVGVEGPPGAGEGKPGARGGPCPTGLP 120
Db 261 GPGGPPGPKSGRPGAPSGKGDGTAKGPGPVGVEGPPGAGEGKPGARGGPCPTGLP 320
QY 121 GPPGERGGGSRGPPGADGVAGPKPAGRGSPGAPGPKGSPGAGRPGEAGLPKAGL 180

```

```

|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 321 GPPGERGGGSRGPPGADGVAGPKPAGRGSPGAPGPKGSPGAGRPGEAGLPKAGL 380
QY 181 GSPGSPGPDGKTGTPPPAGEDGRPPGPPGARGGAGVMGPPGPKGAAGEPKKAGRGVP 240
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 381 GSPGSPGPDGKTGTPPPAGQDGRPPGPPGARGGAGVMGPPGPKGAAGEPKKAGRGVP 440
QY 241 GPPGAVGPDGKGAGAGEGPPGPPGAPGAGERGEEGPAGSPGPEGLPGPAGPPGEGAKPGEE 300
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 441 GPPGAVGPDGKGAGAGEGPPGPPGAPGAGERGEEGPAGSPGPEGLPGPAGPPGEGAKPGEE 500
QY 301 GVPGLGAPGSPGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 501 GVPGLGAPGSPGARGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 560
QY 361 GAPGLEGMPPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGTPTGPPGPPGAPGDK 420
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 561 GAPGLQMPGERGAAGLPGPKGDRGDAGPKGADGSPGKDGVRGLTGTPTGPPGPPGAPGDK 620
QY 421 GESGSPGAPGTGARGAPGDRGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 621 GESGSPGAPGTGARGAPGDRGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 680
QY 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 540
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 681 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 740
QY 541 GKEG 544
Db 741 GKEG 744

RESULT 15
AAY56800
ID AAY56800 standard; protein; 1411 AA.
XX
AC AAY56800;
XX
DT 27-MAR-2000 (first entry)
XX
DE Human preproalpha 1 (I) collagen.
XX
KW Fibrillar collagen; C propeptide; SSAD; telopeptide; gelatin;
KW sequence selection and alignment domain; prosthetic implant; foodstuff;
KW medicine; type I collagen; human.
XX
OS Homo sapiens.
XX
PN EP967226-A2.
XX
PD 29-DEC-1999.
XX
PF 04-MAY-1999; 99EP-00303470.
XX
PR 08-MAY-1998; 98US-0084828P.
PR 10-APR-1999; 99US-00289578.
XX
PA (COHE-) COHESION TECHNOLOGIES INC.
XX
PI Olsen DR, Hitzeman RA, Chisholm GE;
XX
DR WPI; 2000-074666/07.
XX
PT New method for production of fibrillar collagen, useful for preparing
PT telopeptide collagen fibrils and gelatin.
XX
XX Example 1; Fig 3A-B; 30pp; English.
XX
CC The invention provides a method for the production of fibrillar collagen.
CC The method comprises: (a) culturing a recombinant host cell comprising a
CC DNA encoding a fibrillar collagen monomer lacking a C propeptide SSAD
CC (sequence selection and alignment domain); and (b) producing the
CC fibrillar collagen. The methods are used to produce fibrillar collagen,

```

CC from which telopeptide collagen fibrils can be derived. Host cells,
CC comprising DNA encoding a collagen monomer lacking SSAD or N propeptide
CC is used to produce gelatin. Collagen is used in biological research as a
CC substrate for in vitro cell culture and as a component of biocompatible
CC materials for use in prosthetic implants, sustained drug release
CC matrices, artificial skin and wound dressing and healing devices. Gelatin
CC is particularly useful for foodstuffs and medicine, for coating tablets
CC and making capsules. The methods, comprising the use of collagen monomers
CC lacking the N and/or C propeptides, result in a large increase in the
CC production of type I collagen. The present sequence represents the human
CC preproalpha 1 (I) collagen (GenBank Accn no: AF017178)
XX
SQ Sequence 1411 AA;

Query Match 96.6%; Score 2966; DB 3; Length 1411;
Best Local Similarity 95.6%; Pred. No. 3.6e-169;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;
QY 1 GSEGEVGRGEPGPGAGAGADPGADGEPGAKGADGAPGIAGAPGFFGARGPSGPE 60
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
359 GSEGPQVGRGEPGPGAGAGAGADPGADGEPGAKGADGAPGIAGAPGFFGARGPSGPE 418
QY 61 GPGGPPGKGDGEPGAPGSKDGTGAKGPPGVGVGPPGAGGKPGARCEPGPTGLP 120
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
419 GFGGPPGKNGGEPGAPGSKDGTGAKGPPGVGVGPPGAGGKPGARCEPGPTGLP 478
QY 121 GPPGERGGSGRGGFGADGAVGPKGPAERGSFGPAGPKGSPGEAGRPGEAGLPGAKGLT 180
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
479 GPPGERGGSGRGGFGADGAVGPKGPAERGSFGPAGPKGSPGEAGRPGEAGLPGAKGLT 538
QY 181 GSPGSPGPKTGTPGPPAGGEDRPPGPPGARGAGVWGFPGPKGAGEPKAGRGVVP 240
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
539 GSPGSPGPKTGTPGPPAGQDGRPPGPPGARGAGVWGFPGPKGAGEPKAGRGVVP 598
QY 241 GPPGAVGPKDGEAGAGPPGPPGAGRGERGPPGPPGPPGPPGPPGPPGPPGPPGPP 300
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
599 GPPGAVGPKDGEAGAGPPGPPGAGRGERGPPGPPGPPGPPGPPGPPGPPGPPGPP 658
QY 301 GVPGLGAPGSGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 360
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
659 GVPGLGAPGSGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 718
QY 361 GAPGLECMFGERGAAGLPGPKDGRDAGPKGADGSPKDGVRGLTGPTGPPGAPGDK 420
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
719 GAPGLQGMFGERGAAGLPGPKDGRDAGPKGADGSPKDGVRGLTGPTGPPGAPGDK 778
QY 421 GESGSPGAPGTCARGAPDRGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
779 GESGSPGAPGTCARGAPDRGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 838
QY 481 GPAGPAGPPGPIGDVGAPKAGSAGPPGATGFFGAAAGRVGPPGPPGSDAGPPPPGPA 540
DB |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
839 GPAGPAGPPGPIGNVGAPKAGSAGPPGATGFFGAAAGRVGPPGPPGSDAGPPPPGPA 898
QY 541 GKEG 544
DB |||||
899 GREG 902

Search completed: June 17, 2005, 15:08:28
Job time : 90.8873 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:46:04 ; Search time 23.0908 Seconds
(without alignments)
1758.668 Million cell updates/sec

Title: US-10-658-989A-4

Perfect score: 3070

Sequence: 1 GSEGPVGRGPPGPPGAGA.....PPSGDAGPPGPPGACKEG 544

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2966	96.6	1461	4	US-09-585-887-9
2	2966	96.6	1461	4	US-09-289-578-9
3	2966	96.6	1464	4	US-09-331-347C-21
4	2857	93.1	822	3	US-09-219-849-49
5	2842	92.6	1057	3	US-08-931-820-1
6	2791.5	90.9	1341	3	US-08-963-825-18
7	2791.5	90.9	1341	3	US-09-500-811-18
8	2791.5	90.9	1341	3	US-09-570-573-18
9	2791.5	90.9	1341	3	US-09-548-608-18
10	2226	72.5	1017	4	US-08-468-996-10
11	2202	71.7	1060	3	US-08-931-820-3
12	2202	71.7	1418	3	US-08-963-825-20
13	2202	71.7	1418	3	US-09-010-993-1
14	2202	71.7	1418	3	US-09-500-811-20
15	2202	71.7	1418	3	US-09-570-573-20
16	2202	71.7	1418	3	US-09-548-608-20
17	2192	71.4	1442	2	US-08-316-650-12
18	2192	71.4	1442	5	PCR-US95-02251-12
19	2156	70.2	595	3	US-09-219-849-48
20	2156	70.2	595	3	US-09-219-849-50
21	1901	61.9	1057	3	US-08-931-820-4
22	1886	61.4	1024	3	US-08-931-820-2
23	1886	61.4	1366	3	US-08-963-825-19
24	1886	61.4	1366	3	US-09-500-811-19
25	1886	61.4	1366	3	US-09-570-573-19
26	1886	61.4	1366	3	US-09-548-608-19
27	1885	61.4	1366	4	US-09-585-887-10

28	1885	61.4	1366	4	US-09-289-578-10	Sequence 10, Appl
29	1885	61.4	1366	4	US-09-949-016-5882	Sequence 5882, Ap
30	1871	60.9	1078	3	US-08-963-825-21	Sequence 21, Appl
31	1871	60.9	1078	3	US-09-500-811-21	Sequence 21, Appl
32	1871	60.9	1078	3	US-09-570-573-21	Sequence 21, Appl
33	1871	60.9	1078	3	US-08-548-608-21	Sequence 21, Appl
34	1640.5	53.4	1064	1	US-08-642-255-62	Sequence 62, Appl
35	1600	52.1	720	3	US-09-219-849-4	Sequence 4, Appl
36	1600	52.1	777	1	US-08-642-255-53	Sequence 53, Appl
37	1522.5	49.6	492	4	US-08-468-996-12	Sequence 12, Appl
38	1502.5	48.9	1065	1	US-08-642-255-72	Sequence 72, Appl
39	1496.5	48.7	561	1	US-08-642-255-52	Sequence 52, Appl
40	1490.5	48.6	504	3	US-09-219-849-3	Sequence 3, Appl
41	1447	47.1	633	1	US-08-642-255-73	Sequence 73, Appl
42	1447	47.1	1745	4	US-08-795-061-4	Sequence 4, Appl
43	1444	47.0	1806	4	US-09-919-497-56	Sequence 56, Appl
44	1414.5	46.1	960	3	US-09-219-849-5	Sequence 5, Appl
45	1406.5	45.8	829	1	US-08-642-255-132	Sequence 132, App

ALIGNMENTS

RESULT 1
US-09-585-887-9 ; Sequence 9, Application US/09585887
; Patent No. 6413742
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/585.887
; CURRENT FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/289,578
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-585-887-9

Query Match		96.6%	Score 2966;	DB 4;	Length 1461;
Best Local Similarity		95.6%	Pred. No. 1.3e-176;		
Matches 520;		Conservative 21;	Mismatches 3;	Indels 0;	Gaps 0;
QY	1	GSEGPVGRGPPGPPGAGAAGADPCADGEPGAKGADGAPGTAGAPGPGARPGSPGE	60		
DB	359	GSEGPVGRGPPGPPGAGAAGADPCADGEPGAKGADGAPGTAGAPGPGARPGSPGPQ	418		
QY	61	GPGGPPGKGSGBPCAPGSKDTCAGKCPGVGVEGPPGAGEGKPGARCGPTGLP	120		
DB	419	GPGGPPGKGSGBPCAPGSKDTCAGKCPGVGVEGPPGAGEGKPGARCGPTGLP	478		
QY	121	GPBERGPPGSRGPPGADGAPGKPGAGERSGPGAGPKGSPGAGRPCEAGLPCAAGIT	180		
DB	479	GPBERGPPGSRGPPGADGAPGKPGAGERSGPGAGPKGSPGAGRPCEAGLPCAAGIT	538		
QY	181	GSPGSPGDKGTGTPGPPAGEDCRPGPPGARGGAGVMGPPGPKGAAGEPGKAGRGVP	240		
DB	539	GSPGSPGDKGTGTPGPPAGEDCRPGPPGARGGAGVMGPPGPKGAAGEPGKAGRGVP	598		
QY	241	GPPGAVGAGKDGAGAGGPPGPPGAGREGEPPAGSPGEGLPGPAGPPGEAKPGEE	300		

APPLICATION NUMBER: 08/187,319
FILING DATE: 08/187,319
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)
US-09-500-811-18

Query Match 90.9%; Score 2791.5; DB 3; Length 1341;
Best Local Similarity 87.6%; Pred. No. 8.2e-166;
Matches 500; Conservative 21; Mismatches 23; Indels 27; Gaps 4;
QY 1 GSEGEVGRGPPGPPGAGAGPAGDPGADGE---PGAKGADGAPGIAGAPGPGAR--- 54
DB 211 GASGPMGPRGPPGPPGKBGBZAGKPGZRGPPGPGZGARGLPGTAGLPGMKHGRGFS 270
QY 55 -----GPSGPEG-----PG--GPPGPKGDSGPPGAPGSGKDTGAKGPGPV 93
DB 271 GLBGAKGABGAPGAGKZGPGSCZGBCGAPZMGPPGPKGNSGPPGAPSGKDTGAKGPGPV 330
QY 94 GVEGPPGAPGEGKPGARGEPTGLPGPPGERGPPGSGRPPGADGVAGKPGPAGERGSP 153
DB 331 GVQPPGAPGEGKPGARGEPTGLPGPPGERGPPGSGRPPGADGVAGKPGPAGERGSP 390
QY 154 GPAGPKGSPGABRGEAGLPGAKGLTSGPSGPDGKTGTPPGPAGEDGRGPPGPGAR 213
DB 391 GPAGPKGSPGABRGEAGLPGAKGLTSGPSGPDGKTGTPPGPAGQDGRGPPGPGAR 450
QY 214 GEAGVMGPPGKAGGEPKAGRGVPPGPPGAVGPPGAGKDGAGAGPPGPPGAGERGEE 273
DB 451 GQAGVMGPPGKAGGEPKAGRGVPPGPPGAVGPPGAGKDGAGAGPPGPPGAGERGEE 510
QY 274 GPAGSPGEGLPGPAGPPGEGKPGEGVPGDLGAPGSPGARGEPGPPGERGVEGPPGPA 333
DB 511 GPAGSPGEGLPGPAGPPGEGKPGEGVPGDLGAPGSPGARGEPGPPGERGVEGPPGPA 570
QY 334 GPPGADGAPGDDGAKGDAGAPGAPGSEGAPGLEMPGERGAGLPGPKGDRGDAGPKGAD 393
DB 571 GPRGANGAPGNDGAKGDAGAPGAPGSCQAPGLQGMPPGERGAGLPGPKGDRGDAGPKGAD 630
QY 394 GSPGKDGVRGLTGPTGPPGPPGAPGDKGESGPPGAPGTGARGAPGDRGEPGPPGAPGA 453
DB 631 GSPGKDGVRGLTGPTGPPGPPGAPGDKGESGPPGAPGTGARGAPGDRGEPGPPGAPGA 690
QY 454 GPPGADGPPGKAGGEPGADGAGKDPGPPGAPGPPGPIGVGAPGAKGARGSPGPGAT 513
DB 691 GPPGADGPPGKAGGEPGADGAGKDPGPPGAPGPPGPIGVGAPGAKGARGSPGPGAT 750
QY 514 GPPGAGAGRVGPPGPPGSDAGPPGPPGAPGKEG 544
DB 751 GPPGAGAGRVGPPGPPGSDAGPPGPPGAPGKEG 781

RESULT 8

US-09-570-573-18
; Sequence 18, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per

APPLICANT: Bonde, Martin
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/570,573
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 08/187,319
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)
US-09-570-573-18

Query Match 90.9%; Score 2791.5; DB 3; Length 1341;
Best Local Similarity 87.6%; Pred. No. 8.2e-166;
Matches 500; Conservative 21; Mismatches 23; Indels 27; Gaps 4;
QY 1 GSEGEVGRGPPGPPGAGAGPAGDPGADGE---PGAKGADGAPGIAGAPGPGAR--- 54
DB 211 GASGPMGPRGPPGPPGKBGBZAGKPGZRGPPGPGZGARGLPGTAGLPGMKHGRGFS 270
QY 55 -----GPSGPEG-----PG--GPPGPKGDSGPPGAPGSGKDTGAKGPGPV 93
DB 271 GLBGAKGABGAPGAGKZGPGSCZGBCGAPZMGPPGPKGNSGPPGAPSGKDTGAKGPGPV 330
QY 94 GVEGPPGAPGEGKPGARGEPTGLPGPPGERGPPGSGRPPGADGVAGKPGPAGERGSP 153
DB 331 GVQPPGAPGEGKPGARGEPTGLPGPPGERGPPGSGRPPGADGVAGKPGPAGERGSP 390
QY 154 GPAGPKGSPGABRGEAGLPGAKGLTSGPSGPDGKTGTPPGPAGEDGRGPPGPGAR 213
DB 391 GPAGPKGSPGABRGEAGLPGAKGLTSGPSGPDGKTGTPPGPAGQDGRGPPGPGAR 450
QY 214 GEAGVMGPPGKAGGEPKAGRGVPPGPPGAVGPPGAGKDGAGAGPPGPPGAGERGEE 273
DB 451 GQAGVMGPPGKAGGEPKAGRGVPPGPPGAVGPPGAGKDGAGAGPPGPPGAGERGEE 510
QY 274 GPAGSPGEGLPGPAGPPGEGKPGEGVPGDLGAPGSPGARGEPGPPGERGVEGPPGPA 333
DB 511 GPAGSPGEGLPGPAGPPGEGKPGEGVPGDLGAPGSPGARGEPGPPGERGVEGPPGPA 570

STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/963,825
FILING DATE:
CLASSIFICATION: 436
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/187,319
FILING DATE: 21-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN -ALPHA 1 (I1)
US-08-963-825-20

Query Match 71.7%; Score 2202; DB 3; Length 1418;
Best Local Similarity 71.9%; Pred. No. 3.1e-129;
Matches 391; Conservative 40; Mismatches 113; Indels 0

1	QY	GS	G	S	G	P	G	V	R	G	E	P	G	P	G	P	A	G	A	G	P	D	G	E	P	C	A	K	A	G	A	D	G	A	G	P	G	F	G	A	G	P	G	S	G	P	E	60					
315	Db	GP	S	E	A	G	A	Q	R	G	E	P	T	T	G	S	F	G	A	G	S	N	P	T	D	G	I	P	G	A	K	S	A	G	I	A	G	F	G	P	G	R	G	P	D	Q	374						
61	QY	GP	G	S	P	P	K	D	S	E	P	G	A	P	G	S	G	D	T	G	A	K	E	P	C	P	V	C	E	G	P	P	A	G	E	E	K	P	G	A	R	G	E	P	G	T	G	L	P	120			
375	Db	G	A	T	G	P	L	G	P	K	Q	T	G	K	P	G	I	A	G	F	K	G	Q	P	G	E	P	P	A	G	P	A	G	E	E	K	R	G	A	R	G	E	P	G	V	G	P	I	434				
121	QY	GP	P	E	R	G	G	P	S	R	G	F	P	G	A	D	G	V	A	G	K	P	A	G	E	R	G	S	P	C	P	A	G	P	K	S	P	E	A	R	P	G	E	A	G	L	P	G	A	K	L	T	180
435	Db	GP	P	E	R	G	A	P	N	R	G	F	P	G	Q	D	L	A	G	P	K	A	P	E	R	G	P	S	L	A	G	P	K	A	G	D	P	C	P	E	P	G	L	F	G	A	R	L	T	494			
181	QY	GS	P	S	G	P	P	D	K	T	T	P	P	A	G	E	D	G	R	P	P	P	C	P	G	A	R	E	A	G	A	G	W	F	P	C	P	K	A	G	A	R	P	G	K	A	R	G	V	P	240		
495	Db	GR	E	D	A	G	Q	V	K	V	P	S	G	A	P	G	D	G	R	P	P	P	G	P	G	Q	A	R	Q	P	G	W	F	P	C	P	K	A	G	E	P	G	K	A	R	G	E	K	G	L	P	554	
241	QY	GP	P	A	V	G	P	A	K	D	E	A	G	E	P	P	G	A	G	E	R	G	E	S	P	A	G	S	P	C	F	E	G	L	P	C	P	A	G	P	P	G	P	A	G	K	P	E	E	300			
555	Db	G	A	F	L	R	G	L	P	K	D	E	T	C	A	E	G	P	P	A	G	A	R	E	G	E	G	A	G	P	C	S	F	Q	L	P	P	P	P	P	G	G	R	F	G	D	Q	614					
301	QY	GV	P	D	C	D	L	A	G	P	S	G	A	R	E	P	P	P	B	R	G	V	E	G	P	P	A	G	P	D	A	G	P	D	G	A	K	D	A	G	A	P	A	G	P	S	E	360					
615	Db	GV	F	E	A	G	A	P	L	V	P	R	G	E	P	P	G	S	F	G	A	Q	L	G	P	R	L	P	C	T	P	T	D	G	P	K	A	S	G	A	P	P	G	A	Q	674							
361	QY	G	A	P	L	E	M	P	O	E	R	A	A	G	L	P	G	K	D	R	G	A	D	P	K	A	D	G	S	P	K	O	G	V	R	L	T	G	P	I	P	P	G	P	A	G	D	K	420				
675	Db	GP	P	L	Q	M	P	G	E	R	A	G	I	A	G	P	K	D	R	V	E	K	P	E	A	G	K	O	G	R	L	T	G	P	I	P	P	P	A	G	A	N	G	E	734								
421	QY	GS	P	S	G	P	P	A	T	T	G	A	R	G	A	D	R	G	E	P	P	P	P	A	G	F	P	A	G	P	C	A	D	G	E	P	C	A	K	E	P	G	D	A	K	G	D	A	G	P	480		
735	Db	GE	V	P	P	G	P	A	S	A	G	A	R	G	A	P	E	R																																			

Qy		481	GPGAPAGPPPTGDDVAGPAGKARGSGAPP	GCATGTPCGAAGRVGPPSPGDAGPPGP	GA 540
				: :	
D _b		795	GQGSSGAPGGPGTGTGTPKGARGAOGPP	GCATGTPCGAAGRVGPPSNNGNFGPPGP	GS 854
			: :	: :	
Qy		541	GKEG	544	
			:		
D _b		855	GKDG	858	
			:		

RESULT 13

US-09-010-999-1
; Sequence 1, Application US/09010999
; Patent No. 6132976
; GENERAL INFORMATION:
; APPLICANT: Poole, Anthony R.
; APPLICANT: Hollander, Anthony P.
; APPLICANT: Billinghamurst, R. C.
; TITLE OF INVENTION: IMMUNOASSAYS FOR THE MEASUREMENT OF
; TITLE OF INVENTION: COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,999
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 4335
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/448,501
; FILING DATE: 17-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/984,123
; FILING DATE: 04-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 032931/0212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Human Type II Collagen
US-09-010-999-1

Query Match	71.7%;	Score 2202;	DB 3;	Length 1418;
Best Local Similarity	71.9%;	Pred. No. 3.1e-129;		
Matches 391;	Conservative	40;	Mismatches 113;	Indels 0;
			Gaps	0;

[illegible]

COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/570,573
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/187,319
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1418 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN -ALPHA 1 (I1)
US-09-570-573-20

[illegible]

	Query Match	100.00%	Score 3070;	DB 17;	Length 544;
	Best Local Similarity	100.00%	Pred. No. 4e-144;		
	Matches 544;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GSEGGVGGECPGCPAGAACGAGDPGADGEPGKAGDAGCAGCAGCGPFGARGSGPE	60		
Db	1	GSEGGVGGECPGCPGAGAACGAGDPGADGEPGKAGDAGCAGCAGCGPFGARGSGPE	60		
Qy	61	GPGGPPGPKDSGEPGAPGSGKDTGAKGPPGVGVEGPPGAGECKPGARGPPGTGLP	120		

Db	61	GP	GGPGPGPKD	SGEPGAPGSGK	GDTCAGKEBPGVGVG	VEGPPGAPGEGP	KGAPGSGP	GTGLP	120	
Qy	121	GP	PERGEGSGR	GFP	GADGVAGPKGP	PAGERGSPGAPGKSGP	GAGRPG	EAGLPG	AKGLT 180	
Db	121	GP	PERGGPGSR	GFP	GADGVAGPKGP	PAGERGSPGAPGKSGP	GAGRPG	EAGLPG	AKGLT 180	
Qy	181	GS	PGSPGPDGK	TGPP	CGPAGEDGR	PPGPPGARG	EAGVMGPPGPKGA	AGBPGK	AGRGVP 240	
Db	181	GS	PGSPGPDGK	TGPP	CGPAGEDGR	PPGPPGARG	EAGVMGPPGPKGA	AGBPGK	AGRGVP 240	
Qy	241	GPP	CAVGPAGK	DCEAG	AEPPGPPGAPG	AGERBEGP	PAGSPGEG	LPGPAGPPG	AGAPGEE 300	
Db	241	GPP	CAVGPAGK	DCEAG	AEPPGPPGAPG	AGERBEGP	PAGSPGEG	LPGPAGPPG	AGAPGEE 300	
Qy	301	GV	PDGLGAP	PGPSG	ARGEPPG	GERGVEG	PPGPPGADG	APGDDG	AKGDAGAPG	AGSE 360
Db	301	GV	PDGLGAP	PGPSG	ARGEPPG	GERGVEG	PPGPPGADG	APGDDG	AKGDAGAPG	AGSE 360
Qy	361	GAP	LESGMP	GERG	AGLPGPK	CDRGDAGPK	ADSGPK	DGVRGLTG	PIGP	PPGAPGDK 420
Db	361	GAP	LESGMP	GERG	AGLPGPK	CDRGDAGPK	ADSGPK	DGVRGLTG	PIGP	PPGAPGDK 420
Qy	421	G	ESGSPG	PAGT	CARGAP	DGRGEP	PPGPPGADG	EPGAKG	EPDG	AKGDAGPP 480
Db	421	G	ESGSPG	PAGT	CARGAP	DGRGEP	PPGPPGADG	EPGAKG	EPDG	AKGDAGPP 480
Qy	481	GP	APGAPG	PGPT	GDVGA	CPGAKG	AGSGACPP	CGATCF	PGCAAGRV	PPGPPGPPGA 540
Db	481	GP	APGAPG	PGPT	GDVGA	CPGAKG	AGSGACPP	CGATCF	PGCAAGRV	PPGPPGPPGA 540
Qy	541	GK	EX	544						
Db	541	GK	EX	544						

```

RESULT 2
US-10-901-816A-5
; Sequence 5, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901.816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-5

```

QY	181	GS	PSGSPGDKTGP	PPGAGEDGR	PGPPGGPGAR	GEAGVMFP	PKPKAA	GEPC	KAGRGVP	240			
Db	364	GS	PSGSPGDKTGP	PPGAGEDGR	PGPPGGPGAR	GEAGVMFP	PKPKAA	GEPC	KAGRGVP	423			
QY	241	GP	PGAVG	PAGKDGE	AGAGPP	PGPAG	PAGER	GEGPAG	SPGFGL	PCPAGPPGEAGKPGSE	300		
Db	424	GP	PGAVG	PAGKDGE	AGAGPP	PGPAG	PAGER	GEGPAG	SPGFGL	PCPAGPPGEAGKPGQ	483		
QY	301	GV	PDLDG	APGSGAR	GEPCFP	FGERG	VEGPP	GGPAG	PPGADG	APDDGAKD	GAPAGPSE	360	
Db	484	GV	PDLDG	APGSGAR	GEPCFP	FGERG	VQVGP	PGPAG	PRGANG	APGNDGAK	GADGAPGAPSQ	543	
QY	361	GAP	GLEGMP	GERGAAGL	PGPKD	RGDAG	PKGAD	GSPGK	DGVRGL	TGPTGPP	PGPAGAPGDK	420	
Db	544	GAP	GLQGM	PERGAAGL	PGPKD	RGDAG	PKGAD	GSPGK	DGVRGL	TGPTGPP	PGPAGAPGDK	603	
QY	421	GES	PSGSPAG	PTGARG	APGDR	GERG	EPGPP	PGPAG	PCPAG	ADGEP	GKGEPC	DAGAKDAGPP	480
Db	604	GES	PSGSPAG	PTGARG	APGDR	GERG	EPGPP	PGPAG	PCPAG	ADGEP	GKGEPC	DAGAKDAGPP	663
QY	481	GP	AGPAG	PPGPI	GDVGAP	GKAG	SGSAG	PPCAT	CFPGA	RGVPP	GPSGDAG	PPGPPGPA	540
Db	664	GP	AGPAG	PPGPI	IGNVAP	PKAK	SGSAG	PPCAT	CFPGA	RGVPP	GPSGDAG	PPGPPGPA	723
QY	541	GKEG	544										
Db	724	GKEG	727										

RESULT 3
US-10-901-816A-6
; Sequence 6, Application US/10901816A
; Publication No. US20050058703A1

```

; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901.816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 1014
; TYPE: prt
; ORGANISM: Homo sapiens
; US-10-901-816A-6

```

[illegible][illegible]

QY 541 GKEG 544
 Db 741 GKEG 744

RESULT 7
 US-10-104-889-6
 ; Sequence 6, Application US/10104889
 ; Publication No. US20040086961A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GRUSKIN, ELLIOT A.
 ; BROKAW, JANE
 ; ZHANG, GUANGHUI
 ; PAOLELLA, DAVID
 ; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
 ; NUMBER OF SEQUENCES: 50
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DILWORTH & BARRESE
 ; STREET: 333 EARLE O'VINGTON BOULEVARD
 ; CITY: UNIONDALE
 ; STATE: NY
 ; COUNTRY: U.S.A.
 ; ZIP: 11553
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/104,889
 ; FILING DATE: 22-Mar-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/169,768
 ; FILING DATE: 09-OCT-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEEN, JEFFREY S
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (516) 228-8484
 ; TELEFAX: (516) 228-8516
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1169 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 ; US-10-104-889-6

RESULT 8
 US-10-104-889-8
 ; Sequence 8, Application US/10104889
 ; Publication No. US20040086961A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GRUSKIN, ELLIOT A.
 ; BROKAW, JANE
 ; ZHANG, GUANGHUI
 ; PAOLELLA, DAVID
 ; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
 ; NUMBER OF SEQUENCES: 50
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DILWORTH & BARRESE
 ; STREET: 333 EARLE O'VINGTON BOULEVARD
 ; CITY: UNIONDALE
 ; STATE: NY
 ; COUNTRY: U.S.A.
 ; ZIP: 11553
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/104,889
 ; FILING DATE: 22-Mar-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/169,768
 ; FILING DATE: 09-OCT-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: STEEN, JEFFREY S
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (516) 228-8484
 ; TELEFAX: (516) 228-8516
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1171 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
 ; US-10-104-889-8

Query Match 96.6%; Score 2966; DB 15; Length 1171;
 Best Local Similarity 95.6%; Pred. No. 9.6e-139;
 Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

Query Match 96.6%; Score 2966; DB 15; Length 1169;
 Best Local Similarity 95.6%; Pred. No. 9.6e-139;
 Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGEVGRGPPGPPGAGAGPAGDPCADCEPAGKAGADGAPGAGPFGPARGPSGPE 60
 Db 201 GSEGPQVGRGPPGPPGAGAGPAGNPGADGQPGKANGAGPAGPFGARGPSGQ 260

QY 61 GPGGPPGKGSBRPAGPSKGDTCAGKEPGFVGVGPPGAGKEGKPGARGEPGPTGLP 120
 Db 261 GPGGPPGKGSBRPAGPSKGDTCAGKEPGFVGVGPPGAGKEGKPGARGEPGPTGLP 320

QY 121 GPPGCGGSGRGPAGDGVAGKPGAGERSGPGAGKPSGPGAGRGEAGLPCAKGLT 180
 Db 321 GPPGCGGSGRGPAGDGVAGKPGAGERSGPGAGKPSGPGAGRGEAGLPCAKGLT 380

QY 181 GSPGSGPDGKTGPPGAGEGRCRPPGPPGARGAGVGMGPPGKGAAGEGKAGRGVP 240
 Db 381 GSPGSGPDGKTGPPGAGEGRCRPPGPPGARGAGVGMGPPGKGAAGEGKAGRGVP 440

QY 241 GPPGAVGPAGKDEAGAEPPGPPGAPGAGERGEGPAGSPGFEGLFPGAPPPGAEAKPGEE 300

Query Match	96.6%	Score 2966;	DB 14;	Length 1464;
Best Local Similarity	95.6%;	Pred. No. 1.1e-138;		
Matches 520;	Conservative 21;	Mismatches 3;	Indels 0;	Gaps 0;
QY	1	GSEGGEGVRGPPGPPGAGAACGADPCADGCECAKAGADGAPGACGAPGPPGAGAGPSGPE	60	
Db	362	GSEGGQGVRGPPGPPGAGAGPAGNPGADQDQPCAKGANGAPGACGAPGPPGAGPSGPPQ	421	
QY	61	GPGGPPGPKGDSGPPGAPGSGKDTGAKGEGPVPVGVGPPGPPGAPGEGKPGARBPPTGLP	120	
Db	422	GPGGPPGPKGNSGPPGAPGSGKDTGAKGEGPVPVGVGPPGPPGAPGEGKRGARBPPTGLP	481	
QY	121	GPGBRGPGSGRPPGADGVAGPKGAPAGERGSGPPGAPGKSGPCEAGRPEAGLPGAKGLT	180	
Db	482	GPGBRGPGSGRPPGADGVAGPKGAPAGERGSGPPGAPGKSGPCEAGRPEAGLPGAKGLT	541	
QY	181	GSFGSPGDKTGTGPPGAGEDGRPPGPPGARGEAGVMGFPGPKGAAGEPGKAGRGVP	240	
Db	542	GSFGSPGDKTGTGPPGAGQDGRPPGPPGARGQAGVMGFPGPKGAAGEPGKAGRGVP	601	
QY	241	GPBGAVGAGKDGAGAGGPPGAPGAPAGERGEGSPAGSPCFEGLPGPAGPPGEGAKPGEE	300	
Db	602	GPBGAVGAGKDGAGAGGPPGAPGAPAGERGEGSPAGSPGFQGLPGPAGPPGEGAKPGSQ	661	
QY	301	GVPGDLGAPGSGARGEPGPPGEGRVGPPGPPGAPGADGADGCAKGDAGACGAPGSE	360	
Db	662	GVPGDLGAPGSGARGEGFPGEHGVQPPGPPGAPGANGAPGNDGAKGDAGACGAPGSG	721	
QY	361	GAPGLEMPGREGAAGLPGPKDRDAGPKGADGSPGKGVRLGTGTPGPPGAPAGDK	420	
Db	722	GAPGLQMPGREGAAGLPGPKDRDAGPKGADGSPGKGVRLGTGTPGPPGAPAGDK	781	
QY	421	GESGSGPAGPTGARGAGPDGRGPPGAPGAFGPPGADGCEPCAKGCEPCDAGAKGDAGPP	480	
Db	782	GESGSGPAGPTGARGAGPDGRGPPGAPGAFGPPGADGCEPCAKGCEPCDAGAKGDAGPP	841	
QY	481	GPAGPAGPPGPTGDVGAFCAGKARGSGAPCATGPPGAAGRVPPGPPGSDAGPPGPPGA	540	
Db	842	GPAGPAGPPGPTGNVGAFCAGKARGSGAPCATGPPGAAGRVPPGPPGSDAGPPGPPGA	901	
QY	541	GKEG 544		
Db	902	GKEG 905		

```

RESULT 15
US-10-149-352-2
; Sequence 2, Application US/10149352
; Publication No. US20030105050A1
; GENERAL INFORMATION:
; APPLICANT: Beri, Rajinder
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES
; FILE REFERENCE: 06275-254US1
; CURRENT APPLICATION NUMBER: US/10/149,352
; CURRENT FILING DATE: 2002-08-10
; PRIOR APPLICATION NUMBER: PCT/GB00/04741
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: GB 9929487.8
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 4.0
; SEQ ID NO 2
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-149-352-2

Query Match          96.6%; Score 2966; DB 14; Length 1464;
Best Local Similarity 95.6%; Pred. No. 1.1e-138;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps
                                :
Qy      1   GSEGPBGVGRGPPGPGAGAGPAGDPGADGEPGAKGADGAPGIAGAPGPFPGARGPSGPE 60

```

Search completed: June 17, 2005, 15:35:49
Job time : 61.5892 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2005, 13:42:55 ; Search time 18.6216 Seconds
(without alignments)
2810.812 Million cell updates/sec

Title: US-10-658-989A-4
Perfect score: 3070
Sequence: 1 GSEGFGEVGRGEPGPAGG.....PPSGDAGPPGPGACGEG 544

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2966	96.6	1464	1 CGHUI5	collagen alpha 1(I)
2	2857	93.1	1453	2 S21626	collagen alpha 1(I)
3	2775	90.4	1042	1 CGCH15	collagen alpha 1(I)
4	2220	72.3	1418	2 T45467	collagen alpha 1(I)
5	2202	71.7	1487	1 CGHU6C	collagen alpha 1(I)
6	2189	71.3	1419	2 A41182	collagen alpha 1(I)
7	2189	71.3	1487	2 B41182	collagen alpha 1(I)
8	2154	70.2	671	1 CGRT1S	collagen alpha 1(I)
9	2147	69.9	1486	1 B40333	collagen alpha 1(I)
10	2131	69.4	1492	2 A40333	collagen alpha 1(I)
11	2076.5	67.6	779	1 CGHO1S	collagen alpha 1(I)
12	1921.5	62.6	1496	1 CGHU2V	collagen alpha 2(I)
13	1915	62.4	1373	1 A43291	collagen alpha 2(I)
14	1910	62.2	1497	2 I43607	procollagen type V
15	1908	62.1	1464	2 S59856	collagen alpha 1(I)
16	1905.5	62.1	1049	1 CGBO7S	collagen alpha 1(I)
17	1901	61.9	1466	1 CGHU7L	collagen alpha 1(I)
18	1886	61.4	1366	2 I50694	collagen alpha 2(I)
19	1881	61.3	886	2 I50694	collagen alpha 1(I)
20	1550.5	50.5	1414	1 S23809	collagen alpha 1(I)
21	1457.5	47.5	1838	1 CGHU1V	collagen alpha 1(I)
22	1444	47.0	1806	1 CGHU1V	collagen alpha 1(I)
23	1432.5	46.7	1843	2 S18803	collagen alpha 1(I)
24	1418.5	46.2	1027	2 S28774	collagen alpha 1(I)
25	1413	46.0	964	1 CGCH2S	collagen alpha 2(I)
26	1411.5	46.0	1024	2 S18251	collagen alpha 1(I)
27	1400	45.6	888	2 S28791	collagen alpha 1(I)
28	1399.5	45.6	1546	1 CGHU2E	collagen alpha 2(I)
29	1376	44.8	1691	1 S22917	collagen alpha 5(I)

30 1375.5 44.8 2944 2 A54849 collagen alpha 1(I)

31 1372 44.7 1763 2 S16366 collagen alpha 2(I)

32 1364.5 44.4 1549 2 T48103 type VII collagen

33 1357.5 44.2 1690 1 CGHUIB collagen alpha 4(I)

34 1339.5 43.6 1603 2 S23810 collagen alpha 1(I)

35 1339.5 43.6 1670 1 CGHU3B collagen alpha 3(I)

36 1335.5 43.5 920 2 A45748 collagen alpha 1(I)

37 1320 43.0 1669 1 CGHU4B collagen alpha 1(I)

38 1310.5 42.7 1758 2 T29350 hypothetical prote

39 1310 42.7 1669 1 CGMS4B collagen alpha 1(I)

40 1309 42.6 1759 2 T29351 collagen alpha 2(I)

41 1303.5 42.5 812 2 S31521 collagen COLF1 - f

42 1286.5 41.9 1712 1 CGHU2B collagen alpha 2(I)

43 1279.5 41.7 754 2 A55267 collagen alpha 5(I)

44 1277 41.6 921 2 S42617 collagen alpha 1(I)

45 1276.5 41.6 1142 2 JX0369 collagen alpha 1(I)

ALIGNMENTS

RESULT 1

CGHUI5
collagen alpha 1(I) chain precursor - human
N:Alternate names: procollagen alpha 1(I) chain
C:Species: Homo sapiens (man)
C:Date: 12-Aug-1981 #sequence_revision 04-Oct-1996 #text change 09-Jul-2004
C:Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S1135269; A29439; I53466; A02852; I37247
R:D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.
Gene 67, 105-115, 1988
A:Title: Complete nucleotide sequence of the region encompassing the first twenty-five ex
A:Reference number: I60114; MUID:88329734; PMID:2843432
A:Accession: I60114
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-369, 'L', 371-589 <DAL>
A:Cross-references: UNIPROT:P02452; UNIPROT:Q14992; UNIPROT:Q16053; UNIPROT:Q13896; UNIP
R:Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Prock
Biochem. J. 253, 919-922, 1988
A:Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human t
A:Reference number: S01143; MUID:89025644; PMID:3178743
A:Accession: S01143
A:Molecule type: mRNA
A:Residues: 1-472 <TRO>
A:Cross-references: EMBL:X07884; NID:G30015; PIDN:CAA30731.1; PID:G30016; GB:M36546; NID:
A:Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.
Nature 310, 337-340, 1984
A:Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of
A:Reference number: A93335; MUID:84270697; PMID:6462220
A:Accession: A93335
A:Molecule type: DNA
A:Residues: 1-58, 'Q', 60-181 <CHU>
A:Cross-references: EMBL:X00820; NID:G35657; PIDN:CAA25394.1; PID:G35658
R:Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.
J. Biol. Chem. 262, 15151-15157, 1987
A:Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enha
A:Reference number: I55254; MUID:88033098; PMID:2822714
A:Accession: I55254
A:Status: translation not shown; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-45 <ROS>
A:Cross-references: GB:J02829; NID:G180387; PIDN:AAA51993.1; PID:G180388
R:Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gellinas, R.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A:Title: Regulatory elements in the first intron contribute to transcriptional control of
A:Reference number: A39943; MUID:88097389; PMID:3480516
A:Accession: A39943
A:Molecule type: DNA
A:Residues: 1-34 <BOR>
A:Cross-references: GB:J03559; NID:G180876; PIDN:AAA52052.1; PID:G553238
R:Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

J. Biol. Chem. 260, 2315-2320, 1985
 A>Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
 A/Reference number: I55237; MUID:85130970; PMID:2857713
 A/Accession: I55237
 A>Status: translation not shown; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-34 <CH2>
 A/Cross-references: GB:M0627; NID:g180383; PIDN:AAA51992.1; PID:g553226
 R/Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
 J. Biol. Chem. 265, 6312-6317, 1990
 A>Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina
 rome, type VII.
 A/Reference number: A35233; MUID:90202908; PMID:2318855
 A/Accession: A35233
 A/Molecule type: protein
 A/Residues: 33-52 <WIR>
 A/Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
 R/Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
 EMBO J. 8, 1705-1710, 1989
 A>Title: A base substitution in the exon of a collagen gene causes alternative splicing
 A/Reference number: S09400; MUID:89356643; PMID:2767050
 A/Accession: S09400
 A/Molecule type: mRNA
 A/Residues: 156-183 <WEI>
 R/Click, E.M.; Bornstein, P.
 Biochemistry 9, 4699-4706, 1970
 A>Title: Isolation and characterization of the cyanogen bromide peptides from the alpha1
 A/Reference number: A90567; MUID:71038625; PMID:5529814
 A/Contents: CNBR0-1, CNBR2, CNBR4, CNBR5
 A/Accession: B90567
 A/Molecule type: protein
 A/Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'
 A/Experimental source: skin
 A/Note: evidence for 170-allysine
 R/Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, P
 Eur. J. Biochem. 192, 153-159, 1990
 A>Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
 A/Reference number: S11372; MUID:90382436; PMID:2169412
 A/Accession: S11372
 A/Molecule type: protein
 A/Residues: 175-187;274-287, 'P', 289 <BAE>
 A/Note: sequence of collagen alpha 1(S)(I) isolated from bone after pepsin digestion
 R/Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
 J. Biol. Chem. 266, 21827-21832, 1991
 A>Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
 cooperative melting of intact type I collagen.
 A/Reference number: I55342; MUID:92042092; PMID:1718984
 A/Accession: I55342
 A>Status: translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 258-268;1347-1357 <DEA>
 A/Cross-references: GB:S67495; NID:g239007; PIDN:AAB20350.1; PID:g239008
 A/Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
 R/Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
 J. Biol. Chem. 245, 5042-5048, 1970
 A>Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
 A/Reference number: A92069; MUID:71001508; PMID:4319110
 A/Accession: A92069
 A/Molecule type: protein
 A/Residues: 263-268 <MOR>
 A/Experimental source: skin
 A/Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
 R/Labhard, M.E.; Hollister, D.W.
 Matrix 10, 124-130, 1990
 A>Title: Segmental amplification of the entire helical and telopeptide regions of the cd
 A/Reference number: S15989; MUID:90326017; PMID:2374517
 A/Accession: S15989
 A/Molecule type: mRNA
 A/Residues: 281-302;402-420;823-843;925-944;1026-1045;1143-1162 <LAB>
 R/Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N
 Connect. Tissue Res. 29, 1-11, 1993
 A>Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
 A/Reference number: I52905; MUID:93339042; PMID:8339541

A/Accession: I52905
 A>Status: translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 342-352, 'C', 354-359 <W12>
 A/Cross-references: GB:S64717; NID:g408195; PIDN:AAB27677.1; PID:g408196
 A/Note: mutant sequence from patient with osteogenesis imperfecta
 R/Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
 Biochemistry 22, 5213-5223, 1983
 A>Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha 1 c
 A/Reference number: A30476; MUID:84080385; PMID:6683127
 A/Accession: A30476
 A/Molecule type: mRNA
 A/Residues: 425-1250, 'X', 1252-1328, 'S', 1330-1390, 'X', 1392-1464 <BER>
 A/Cross-references: GB:K0128; NID:g180391; PIDN:AAA51995.1; PID:g180392
 A/Note: sequence partially completed for missing nucleotides by A29439
 R/Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
 J. Biol. Chem. 260, 691-694, 1985
 A>Title: Multixon deletion in an osteogenesis imperfecta variant with increased type III
 A/Reference number: A22161; MUID:85104934; PMID:2981843
 A/Accession: A22161
 A/Molecule type: DNA
 A/Residues: 472-594, 'R', 596-607 <CH3>
 A/Cross-references: GB:K03178; GB:K03179; NID:g179612; NID:g179613; PIDN:AAA51847.1; PID:
 A/Note: the authors translated the codon CGT for residue 595 as Pro
 R/Wallis, G.A.; Starman, B.J.; Zimm, A.B.; Byers, P.H.
 Am. J. Hum. Genet. 46, 1034-1040, 1990
 A>Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained
 A/Reference number: A35336; MUID:90252792; PMID:2339700
 A/Accession: A35336
 A/Molecule type: mRNA
 A/Residues: 710-720, 'B', 722-737, 'E', 739-745 <WAL>
 A/Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
 R/Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes,
 Hum. Mol. Genet. 3, 2201-2206, 1994
 A>Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the ce
 A/Reference number: I54365; MUID:95187161; PMID:7881420
 A/Accession: I54365
 A>Status: translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 746-766, 'S', 768-781 <FOR>
 A/Cross-references: GB:L47667; NID:g1009093; PIDN:AAB59576.1; PID:g1009094
 R/Chesler, S.D.; Wallis, G.A.; Byers, P.H.
 J. Biol. Chem. 268, 18218-18225, 1993
 A>Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of t
 A/Reference number: A47426; MUID:93352646; PMID:8349697
 A/Accession: A47426
 A/Molecule type: mRNA
 A/Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CHE>
 A/Cross-references: GB:S64596; NID:g407589; PIDN:AAB27856.1; PID:g407590
 A/Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBIPI:136445)
 A/Note: does not represent an experimentally determined sequence but three different muta
 A/Accession: B47426
 A/Molecule type: mRNA
 A/Residues: 1179-1464 <CH4>
 A/Experimental source: normal dermal fibroblast culture
 A/Accession: C47426
 A/Molecule type: mRNA
 A/Residues: 1179-1276, 'H', 1278-1464 <CH5>
 A/Experimental source: fetal cell 86-237
 A/Accession: D47426
 A/Molecule type: mRNA
 A/Residues: 1179-1336, 1339-1464 <CH6>
 A/Experimental source: fetal cell 86-146
 A/Accession: E47426
 A/Molecule type: mRNA
 A/Residues: 1179-1387, 'R', 1389-1464 <CH7>
 A/Experimental source: fetal cell 88-251
 R/Cohn, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andressen, P.; Charbonneau, H.; Nid
 J. Biol. Chem. 263, 14605-14607, 1988
 A>Title: Substitution of Cysteine for Glycine within the Carboxyl-terminal Telopeptide of
 A/Reference number: I55269; MUID:89008319; PMID:3170557
 A/Accession: I55269

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1187-1194, 'C', 1196-1220 <OH>

A:Cross-references: GB:M23213; NID:G340842; PIDN:AA859363.1; PID:G499622

A:Note: mutant sequence from a patient with mild osteogenesis imperfecta

R:Maekelae, J.K.; Raasina, M.; Virta, A.; Vuorio, E.

Nucleic Acids Res. 16, 349, 1988

A:Title: Human pro-alpha-1(I) collagen: cDNA sequence for the C-propeptide domain.

Query Match 96.6%; Score 2966; DB 1; Length 1464;
Best Local Similarity 95.6%; Pred. No. 1.1e-135;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GSEGEVGRGPGGPPAGAGPAGDPCADGEPGAKGAGTAGAGPFGCARGPSGPE 60

Db 362 GSEGGVGRGPGGPPAGAGPAGDPCADGEPGAKGAGTAGAGPFGCARGPSGQ 421

Qy 61 GPGGPPGKGSGERGAGPAGSGDTCAGKEGPGVGVGPPGAGBEGKPGCARGEPGPTGLP 120

Db 422 GPGGPPGKGSGERGAGPAGSGDTCAGKEGPGVGVGPPGAGBEGKPGCARGEPGPTGLP 481

Qy 121 GPPGERGGSGRPPGADGAGPAGKGPAGERSGPGAGPKGSPGEGAGRPGEAGLPKAGLT 180

Db 482 GPPGERGGSGRPPGADGAGPAGKGPAGERSGPGAGPKGSPGEGAGRPGEAGLPKAGLT 541

Qy 181 GSPGSGPDGKTGPPGAGEDRPGPPGARGAGVGMGPPGKGAAGEFGKAGRGVP 240

Db 542 GSPGSGPDGKTGPPGAGEDRPGPPGARGAGVGMGPPGKGAAGEFGKAGRGVP 601

Qy 241 GPPGAGVAGKDGAGAGGPPGAGPAGERGEGPAGSPGEGPGLPGPAGPGEAGKPGEE 300

Db 602 GPPGAGVAGKDGAGAGGPPGAGPAGERGEGPAGSPGEGPGLPGPAGPGEAGKPGEE 661

Qy 301 GVPGLGAPGSGARGEFGFERGVEGPPGAGPAGPAGDAGPDDGAKGDAGAPGSGE 360

Db 662 GVPGLGAPGSGARGEFGFERGVEGPPGAGPAGPAGDAGPDDGAKGDAGAPGSGQ 721

Qy 361 GAPGLEHMPGRRGAAGLPKGDAGPKADGSPGKDGVRGLTGPPTGPPGAPGADGX 420

Db 722 GAPGLQMPGRRGAAGLPKGDAGPKADGSPGKDGVRGLTGPPTGPPGAPGADGX 781

Qy 421 GESGSPAGTGTARGAGDGEPPGPPGAGPAGPAGDGPAGKGEPPGAGAGKGDAGPP 480

Db 782 GESGSPAGTGTARGAGDGEPPGPPGAGPAGPAGDGPAGKGEPPGAGAGKGDAGPP 841

Qy 481 GPAGPAGPPTGVDGAPGAKGARGAGPAGTGPAGAGRVGPPGSDAGPAGPPGPA 540

Db 842 GPAGPAGPPTGVDGAPGAKGARGAGPAGTGPAGAGRVGPPGSDAGPAGPPGPA 901

Qy 541 GKEG 544

Db 902 GKEG 905

RESULT 2

S21626

collagen alpha 1(I) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text change 09-Jul-2004

C:Accession: S57243; S16374; A23982; I49557; I49557; I48300; S21626

R:Li, S.W.; Khillan, J.; Prockop, D.J.

Matrix Biol. 14, 593-595, 1994

A:Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I

A:Reference number: S57243

A:Accession: S57243

A:Molecule type: mRNA

A:Residues: 1-1453 <LIS>

A:Cross-references: UNIPROT:P11087; EMBL:U08020; NID:G470673; PIDN:AAA88912.1; PID:G4706

R:Metzsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

Biochim. Biophys. Acta 1089, 241-243, 1991

A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.

A:Reference number: S16176; MUID:91274355; PMID:2054384

A:Accession: S16374

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1442-1453 <NET>

A:Cross-references: EMBL:X57981; NID:G50484; PIDN:CAA1046.1; PID:G50485

R:French, B.T.; Lee, W.H.; Maul, G.G.

Gene 39, 311-312, 1985

A:Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.

A:Reference number: A23982; MUID:86137403; PMID:3841523

A:Accession: A23982

A:Molecule type: mRNA

A:Residues: 518-1128 <PRE>

A:Cross-references: GB:M14423; NID:G192261; PIDN:AAA37333.1; PID:G192262

R:Monson, J.M.; Friedman, J.; McCarthy, B.J.

Mol. Cell. Biol. 2, 1362-1371, 1982

A:Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for a

A:Reference number: I49559; MUID:83141374; PMID:6298597

A:Accession: I49559

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 735-1130 <RES>

A:Cross-references: GB:M17491; NID:G192263; PIDN:AAA37334.1; PID:G192264

R:Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.

Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984

A:Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads to

A:Reference number: I49557; MUID:84170331; PMID:6324198

A:Accession: I49557

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-25 <RES>

A:Cross-references: GB:X01688; NID:G192246; PIDN:AAA37330.1; PID:G553881

R:Fenton, S.P.; Lamande, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.

Biochim. Biophys. Acta 1216, 469-474, 1993

A:Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.

A:Reference number: S39789; MUID:94092741; PMID:8268229

A:Accession: S39789

A:Molecule type: DNA

A:Residues: 1-80, 'E', 82-105, 'D', 107-185; 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-1

R:Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindel, M.

Mol. Cell. Biol. 14, 5950-5960, 1994

A:Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indirect

A:Reference number: I48300; MUID:94344105; PMID:8065328

A:Accession: I48300

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>

A:Cross-references: EMBL:X54876; NID:G50486; PIDN:CAA38657.1; PID:G50487

C:Genetics:

A:Gene: COL1A1

A:Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/3

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C:Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple helix

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-151/Domain: amino-terminal propeptide #status predicted <PRO>

F:30-89/Domain: von Willebrand factor type C repeat homology <WVC>

F:152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>

F:1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 93.1%; Score 2857; DB 2; Length 1453;

Best Local Similarity 91.4%; Pred. No. 1.9e-130;

Matches 497; Conservative 28; Mismatches 19; Indels 0; Gaps 0;

Qy 1 GSEGEVGRGPGGPPAGAGPAGDPCADGEPGAKGAGTAGAGPFGCARGPSGPE 60

Db 351 GSEGGVGRGPGGPPAGAGPAGDPCADGEPGAKGAGTAGAGPFGCARGPSGQ 410

Qy 61 GPGGPPGKGSGERGAGPAGSGDTCAGKEGPGVGVGPPGAGBEGKPGCARGEPGPTGLP 120

Db 411 GPGGPPGKGSGERGAGPAGSGDTCAGKEGPGVGVGPPGAGBEGKPGCARGEPGPTGLP 470

Qy 121 GPPGERGGSGRPPGADGAGPAGKGPAGERSGPGAGPKGSPGEGAGRPGEAGLPKAGLT 180

Db 471 GPPGERGGSGRPPGADGAGPAGKGPAGERSGPGAGPKGSPGEGAGRPGEAGLPKAGLT 530

A;Residues: 752-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'Q', 1038-1052, 'E', 1054-1068, 'T', 'A';
A;Cross-references: GB:L00977; NID:g180812; PIDN:AA23914.1; PID:g258774
A;Note: sequence extracted from NCBI backbone (NCBIP:117273); parts of this sequence were
A;Note: this translation is not annotated and this publication is not cited in GenBank
A;Note: mutant sequence associated with perinatal lethal hypochondrogenesis
R;Tiller, G.E.; Rimoen, D.L.; Murray, L.W.; Cohn, D.H.
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
A;Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual
A;Reference number: S16502; MUID:90251662; PMID:2339128
A;Accession: S16502
A;Molecule type: DNA
A;Residues: 1164-1184, 'GPSKDGANGTGPPI', 1185-1199, 'TII2>
A;Cross-references: EMBL:M37126; NID:g180808; PIDN:AAA52037.1; PID:g180809
A;Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
R;Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
A;Title: Identification and characterization of the human type II collagen gene (COL2A1)
A;Reference number: A02858; MUID:85190534; PMID:3857598
A;Accession: A02858
A;Molecule type: DNA
A;Residues: 1032-1056, 'N', 1058-1068, 'T', 1070-1487, 'CHE>
A;Cross-references: GB:J00116; NID:g180395; PIDN:AAA51997.1; PID:g180396
R;Elima, K.; Vuorio, T.; Vuorio, E.
Nucleic Acids Res. 15, 9499-9504, 1987
A;Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) c
A;Reference number: A27280; MUID:88067771; PMID:2825137
A;Accession: A27280
A;Molecule type: DNA; mRNA
A;Residues: 1175-1487, 'ELI>
A;Cross-references: EMBL:X06268; NID:g30096; PIDN:CAA29604.1; PID:g30097
A;Experimental source: fetal epiphyseal cartilage
R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
Biochem. J. 237, 923-925, 1986
A;Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
A;Reference number: A57033; MUID:87099927; PMID:3800925
A;Accession: A57033
A;Molecule type: protein
A;Residues: 'XE', 1244-1246, 'N', 1248, 'X', 1250-1265, 1295-1305, 1395-1408, 'VAN>
A;Note: chondrocalcin identified as released collagen I(II) chain carboxyl-terminal prop
R;Strom, C.M.; Upholt, W.B.
Nucleic Acids Res. 12, 1025-1038, 1984
A;Title: Isolation and characterization of genomic clones corresponding to the human ty
A;Reference number: A21733; MUID:84118798; PMID:6320112
A;Accession: A21733
A;Molecule type: DNA
A;Residues: 1245-1295, 'STR1>
A;Cross-references: EMBL:X00339; EMBL:X00298; NID:g394699; PIDN:CAA25092.1; PID:g4378975
A;Accession: B21733
A;Molecule type: DNA
A;Residues: 894-909, 'PE', 'STR2>
R;Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A;Title: Isolation and partial characterization of genomic clones coding for a human pro
gene.
A;Reference number: A24561; MUID:86104139; PMID:3002437
A;Accession: A24561
A;Molecule type: DNA
A;Residues: 1296-1358, 'NUN2>
A;Cross-references: GB:M12048; NID:g180017
A;Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
R;Sangliorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez,
Nucleic Acids Res. 13, 2207-2225, 1985
A;Title: Isolation and partial characterization of the entire human pro alpha 1(II) coll
A;Reference number: I37249; MUID:85215609; PMID:2987845
A;Accession: S59491
A;Molecule type: DNA
A;Residues: 7-28, 'R', 99-114, 541-578, 786-802, 1055-1056, 'N', 1058-1068, 'T', 1070-1109, 1200-1
A;Accession: I84453
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 7-28, 'SAN2>

A;Cross-references: GB:M23759; NID:g180845; EMBL:X03320; GB:M24938; NID:g30104
A;Note: the GenBank PID is based on an incorrect reading frame
A;Accession: I37250
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 541-560, 'SAN3>
A;Cross-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929621
A;Accession: I37251

Query Match 71.7%; Score 2202; DB 1; Length 1487;
Best Local Similarity 71.9%; Pred. No. 4.9e-99;
Matches 391; Conservative 40; Mismatches 113; Indels 0; Gaps 0;

QY 1 GSEGGVGRGPPGPPGAGACGADGCPGAGKAGADGAGTAGAGFFGARGPSPE 60
DB 384 GPEGAGRRGPPGTPGSPGAGANPGTDGIPGAKGSAGAGTAGAGFFGPPGPPDQ 443
QY 61 GFGGPPGKSGSGEPGKSGDTCAGKEPVGVEGPPGAGEGKPGARGEPTGLP 120
DB 444 CATGPLGPKGTGKPGIAGFKGEQPKGEPGAGPQGNAGGEGKKGARGEFGVGPI 503
QY 121 GPPGRCGSGRPGADGAVGPKGAGERGSPGAGPKGSPGAGRPGEAGLPKAGLIT 180
DB 504 GPPGSGAPGNRGFFGQDGLAGPKGAPGERGSPGLAGPKGANGDPGRDPEGLPCARGLT 563
QY 181 GSPGSPGDKGTGPPGAGEDRGPDPGPPGARGGAVGMPGPKGAAGERPKAGERV 240
DB 564 GRPDAGQGVKVGSPGAGEDRGPDPGQARQGVGMPGPKGANGEPKAGEGLP 623
QY 241 GPPGAVGAGKDGAGAGPPGPPGAGGERGEGPAGSPGEGPLGPPAGPGEAGKPGEE 300
DB 624 GAPGLRGLPKDGTGAEPPGPPGAPGAGERGEGQAGPSPGQGLGPPGPPGEGKPGDQ 683
QY 301 GVPDGLGAPGSGARGSPGPPGRCVGGPPGPPGADGAPGDDGAKGDAGAPGPGSE 360
DB 684 GVPGEAGAPGLVGRGERGPPGERSGFCAGLQGRGLPGTGTGPKGASGAPGPPCAQ 743
QY 361 GAPGLEGMPPGERRAAGLPGPKDGRDAGPKGADGSPKDGVRGLTGTGTPGPPGAGAPGDK 420
DB 744 GPPGLQGMPPGERRAAGLPGKDRDVGKGPAGPKDGRGLTGTGTPGPPGAGANGK 803
QY 421 GESPPGPPGTPGARGAGRGPDPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
DB 804 GEVPPGPPGAGAGAPGERGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 863
QY 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 540
DB 864 GPQPSGAPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 923
QY 541 GKEG 544
DB 924 GKDG 927

RESULT 6

A4182
collagen alpha 1(II) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C;Accession: A41182; A44885
R;Metsaeranta, M.; Iman, D.; de Crombrughe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
A;Reference number: A41182; MUID:91358469; PMID:1885613
A;Accession: A41182
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-1419, <NET>
A;Cross-references: GB:M65161
R;Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
Development 111, 945-953, 1991
A;Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilage
A;Reference number: A44885; MUID:91347939; PMID:1879363

J. Biol. Chem. 242, 2572-2574, 1967

A;Title: The incomplete hydroxylation of individual prolyl residues in collagen.

A;Reference number: A92029; MUID:67165368; PMID:4230711

A;Contents: CNB8r

A;Accession: A92029

A:Molecule type: protein

A;Residues: 20-55 <B02>

A;Experimental source: skin and tendon

R;Butler, W.T.; Ponds, S.L.

Biochemistry 10, 2076-2081, 1971

A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino acid sequence of the hydroxylated alpha 1(I) chain

A;Reference number: A90353; MUID:71263178; PMID:43273399

A;Contents: CNB8r

A;Accession: A90353

A:Molecule type: protein

A;Residues: 56-102 <BU1>

A;Experimental source: skin

R;Butler, W.T.

Biochemistry 9, 44-50, 1970

A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. The complete amino acid sequence of the alpha 1(I) chain

A;Reference number: A90366; MUID:70085124; PMID:5411206

A;Contents: CNB8r

A;Accession: A90366

A:Molecule type: protein

A;Residues: 103-139 <BU2>

A;Experimental source: skin

R;Balian, G.; Click, E.M.; Bornstein, P.

Biochemistry 10, 4470-4478, 1971

A;Title: Structure of rat skin collagen alphas-CB8. Amino acid sequence of the hydroxylated alpha 1(I) chain

A;Reference number: A90357; MUID:72136131; PMID:4335087

A;Contents: CNB8r

A;Accession: A90357

A:Molecule type: protein

A;Residues: 140-238 <BA1>

A;Experimental source: skin

R;Balian, G.; Click, E.M.; Hermodson, M.A.; Bornstein, P.

Biochemistry 11, 3798-3806, 1972

A;Title: Structure of rat skin collagen alphas-CB8. Amino acid sequence of the hydroxylated alpha 1(I) chain

A;Reference number: A90362; MUID:73006942; PMID:4342027

A;Contents: CNB8r

A;Accession: A90362

A:Molecule type: protein

A;Residues: 239-418 <BA2>

A;Experimental source: skin

R;Butler, W.T.; Underwood, S.P.; Finch Jr., J.E.

Biochemistry 13, 2946-2953, 1974

A;Title: Chemical studies on the cyanogen bromide peptides of rat skin collagen. Amino acid sequence of the alpha 1(I) chain

A;Reference number: A90379; MUID:74271984; PMID:4366532

A;Contents: CNB8r

A;Accession: A90379

A:Molecule type: protein

A;Residues: 419-567 <BU3>

A;Experimental source: skin

R;Stoltz, M.; Timpl, R.; Furthmayr, H.; Kuehn, K.

Eur. J. Biochem. 37, 287-294, 1973

A;Title: Structural and immunogenic properties of a major antigenic determinant in neutral salt-extracted rat skin collagen

A;Reference number: A91209; MUID:74011954; PMID:4126850

A;Contents: CNB8r

A;Accession: A91209

A:Molecule type: protein

A;Residues: 568-651 <ST1>

A;Experimental source: skin

A;Note: This region probably corresponds to positions 949-1032 of the alpha 1(I) chain

R;Stoltz, M.; Timpl, R.; Kuehn, K.

FEBS Lett. 26, 61-65, 1972

A;Title: Non-helical regions in rat collagen alpha-chain.

A;Reference number: A91385; MUID:73049495; PMID:4636751

A;Contents: CNB8r

A;Accession: A91385

A:Molecule type: protein

A;Residues: 651-671 <ST2>

A;Experimental source: skin

A>Note: The composition of peptides comprising residues 1-9 and 1-19 confirms the sequence.
A>Note: This region (residues 651-671 above) probably corresponds to positions 1032-1052
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
ed and subsequently O-glycosylated.
C:Comment: The order of the nine CNBR peptides in the alpha 1(I) chain of rat skin collag
C:Comment: The complete chain contains 1052 residues.
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxyl;
F:I/Modified site: blocked amino end (Glx) (probably pyrrolidone carboxylic acid) #status
F:9/Modified site: allysine (Lys) #status experimental
F:I103/Modified site: carbonylhydrazide (Lys) (covalent) #status experimental
F:I103/Modified site: 5-hydroxylysine (Lys) #status experimental
F:I424,547/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Query Match 70.2%; Score 2154; DB 1; Length 671;
Best Local Similarity 77.3%; Pred. No. 5.5e-97;
Matches 381; Conservative 28; Mismatches 36; Indels 48; Gaps 2;

Qy 1 GSEGEVGRGEGPPGPAGAGPAGDGPAGEPCAKGNADGAPGTAGAPFPFGARGPSPE 60
Db 200 GSEGPGVRGEPGGPPAGAAAGPNAGNFCADGQPCKANGANGAPTAGAPFFGARGPSQG 259

Qy 61 GPQPPPKKDSCEPGAPSGSKGDTGAKEGEPGVGEPPGPAGEEKGFGARGECPTGLP 120
Db 260 GPSGAPKNSCEPGAPNGKDDTGAKGEPGACVQQPFGPAGEEKKRGARGEPGSGLP 319

Qy 121 GPGERGERSRFPAGDAGVAGPKPAGERGSPGPAPGKPSGPAERPGAEAGLPKAIGLT 180
Db 320 GPPERGGSRRFPAGDAGVAGPKPAGERGSPGPAPGKPSGPAERPGAEAGLPKAIGLT 379

Qy 181 GSPGSPGDCKTTPPGPAGSDGDPGPPGPGPARGEAGVMGPPGPKAAGEPKAGERGVP 240
Db 380 GSPGSPGDCKTTPPGPAGZBGRPGPAGPCAQAGVMGFPGPKGTAGEFKAGERGVP 439

Qy 241 GPPCAVGPAGKDRAHAGBPMPAGPAGERGEEGAPGPFEGPLPGPAGPPGEAKGPEG 300
Db 440 GPPCAVGPAGKDRAHAGBPMPAGPAGERGEEGAPGPFEGPLPGPAGPPGEAKGPEGZ 499

Qy 301 GVFDLCAPGPSGARGSPGPFGRGVGPPGPPAGPCADGAPDGDAKGDAGAPGASG 360
Db 500 GVFDLCAPGPSGARGSPGPFGRGVGPPGPPAGPRNNGAPBBGAKGDTGALCAPQSQ 559

Qy 361 GAPLEGMPGERGAAGLPKPDKRDGADPKGADGSPKDGVRGLTGPIPPPAGAPGDK 420
Db 560 GAPDLGWMSGLZ-----GPPSAGSPKBLGLPGLPGIPPPGRTGBAGPS 631

Qy 421 GESGSPGAPTGCARGAPDRGRGPPGPPAGFAGPPGADGEPGAKGFBGADAGDAPP 480
Db 584 GPASGAPAGR-----GPPSAGSPKBLGLPGLPGIPPPGRTGBAGPS 631

Qy 481 GPAGPAGPPGFIG 493
Db 632 GPPGPPGPPGPG 644

RESULT 9
B40333
collagen alpha 1(II) chain precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: B40333
R:Suzuki, H.R.; Bieler, J.J.; Solorush, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A>Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A:Reference number: A40333; PMID:92011898; PMID:1918153
A:Accession: B40333
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1486 <SUA>
A:Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; GB:M63595
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:37-96/Domain: von Willebrand factor type C repeat homology <VMC>

A;Molecule type: protein
A;Residues: 146-294 <FI2>

RESULTS II
CGB01S

A;Residues: 146-294 <FI2>

A:Reference number: A25374; MUID:85289337; PMID:2411731
A:Accession: A25374
A:Molecule type: mRNA
A:Residues: 1227-1417, 'T', 1419-1437, 'S', 1439-1496 <MYE>
A:Cross-references: GB:M1178; NID:g180912; PIDN:AAA52058.1; PID:g180913
A:Experimental source: normal fibroblasts
R:Tsipouras, P.; Schwartz, R.C.; Liddell, A.C.; Salkeld, C.S.; Weil, D.; Ramirez, F.
Genomics 3, 275-277, 1988
A>Title: Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2, located on
A:Reference number: A30017; MUID:89138450; PMID:3224983
A:Accession: A30017
A:Molecule type: DNA
A:Residues: 1449-1463, 'E', 1465-1495, 'A' <TSS>
A:Cross-references: GB:J03051; NID:g179695; PIDN:AAA51958.1; PID:g179696
A>Note: The authors translated the codon GAA for residue 1460 as Gln, and GAG for residue
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
are 5-hydroxylated and subsequently O-glycosylated.
C:Comment: The amino-terminal propeptide domain appears not to be completely cleaved.
C:Genetics:
A:Gene: GDB:COL5A2
A:Cross-references: GDB:119064; OMIM:120190
A:Map position: 2q31-q31
A:Introns: 33/1; 812/3; 830/3; 848/3; 902/3; 922/3; 974/3; 1046/3; 1064/3; 1448/3
A:Complex: type V collagen may be a homotrimer of alpha 1(V) chains (see PIR:CGHUIV), a
alpha 2(V) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the
length, is formed with desmosine cross-links made from lysine and allysine residues
C:Function:
A:Description: structural component of extracellular fibrous polymer associated with cell
A>Note: may play a role in controlling the lateral growth of collagen I fibrils
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1250/Product: collagen alpha 2(V) chain #status predicted <MAT>
F:27-193/Domain: amino-terminal propeptide (uncleaved) #status predicted <NPP>
F:27-108/Region: nonhelical
F:40-99/Domain: von Willebrand factor type C repeat homology <VWC>
F:109-186/Region: helical
F:187-208/Region: nonhelical
F:209-1225/Region: helical
F:503-505/Region: cell attachment (R-G-D) motif
F:941-943/Region: cell attachment (R-G-D) motif
F:1064-1066/Region: cell attachment (R-G-D) motif
F:1067-1069/Region: cell attachment (R-G-D) motif
F:1097-1099/Region: cell attachment (R-G-D) motif
F:1124-1126/Region: cell attachment (R-G-D) motif
F:1133-1135/Region: cell attachment (R-G-D) motif
F:1225-1250/Region: carboxyl-terminal nonhelical telopeptide
F:1251-1496/Domain: carboxyl-terminal propeptide #status predicted <CPP>
F:1269-1496/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F:27/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:193-194/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted
F:194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:201/Modified site: allysine (Lys) #status predicted
F:290,293,296,608,614,1004,1007,1013,1028,1034/Modified site: 4-hydroxyproline (Pro) #st
F:299,1139/Modified site: 5-hydroxylysine (Lys) #status predicted
F:299,1139/Binding site: carboxydrate (Lys) (covalent) #status predicted
F:1025/Modified site: 5-hydroxylysine (Lys) #status experimental
F:1250-1251/Cleavage site: Glu-Asp (procollagen C-endopeptidase) #status predicted
F:1259,1397/Binding site: carboxydrate (Asn) (covalent) #status predicted
F:1293,1299,1325/Disulfide bonds: interchain #status predicted
F:1333-1494,1402-1447/Disulfide bonds: #status predicted

Query Match 62.6%; Score 1921.5; DB 1; Length 1496;
Best Local Similarity 64.0%; Pred. No. 1.4e-85;
Matches 348; Conservative 40; Mismatches 153; Indels 3; Gaps 1;
QY 1 GSEGPVGRGPPPPGAGAGPAGDPGADGPFAGKAGADGAPGAGPFGPAGPFGPAGPFGP 60
DB 396 GPEGPQGRGTGPPGPGVPSGLPGAIGATGDTGPKGPTGSGP---PGSAGPPGSP 452
QY 61 GPGGPPGKDGSGPGAPGSKDGTAKGPGVGVGPPGPPGAGRGKPGARGEPTGLP 120
DB 453 GPQSGTGPQNSGLFGDPGKGEAGPKGSPGHGIGPTIGPPGEGKRGKPRGDPGLTGLP 512

QY 121 GPPGCGPSRSGPPGADGAGVAGKPGPAGRGSPGAPGKSPGEGAGLPGAKGLT 180
DB 513 GPGVGERGAPNRRGPPGSDGLPGPKAQGRGPGVSSGPKSGQDPGRPPGLPGAKGLT 572
QY 181 GSPGSPGDPGKTGTPGAGEDGRPPGPPGARGGAGVWGPPGPKGAAGEPGKAGRGVP 240
DB 573 GNPVGQFEGKGLPLGAPGEDGRPPGSGIGIKQPGTGMGLPGPKGSGNDPKGPGAGNP 632
QY 241 GPPGAVGAGKDGAGAGPPGPPGAPGAGERGEGPAGSPGPEGLPGPAGPPGEGAKGPEE 300
DB 633 GVPQGRGAPKDGKVGPPYPPGPPGLRGERGEGQPPGPTGFGQHPGPPGPPGEGKPGDQ 692
QY 301 GVPGLGAPGSPGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
DB 693 GVPGPPGAVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 752
QY 361 GAPGLEGMPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 420
DB 753 GPPGLQMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 812
QY 421 GESGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
DB 813 GEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 872
QY 481 GPAGPAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 540
DB 873 GPQGLAGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 932
QY 541 GKEG 544
DB 933 GKEG 936
RESULT 13
A43291
collagen alpha 2(I) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: A43291; A54328
R:Phillips, C.L.; Morgan, A.L.; Lever, L.W.; Wenstrup, R.J.
Genomics 13, 1345-1346, 1992
A>Title: Sequence analysis of a full-length cDNA for the murine pro alpha 2(I) collagen
A:Reference number: A43291; MUID:92372043; PMID:1505972
A:Accession: A43291
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1373 <PHI>
A:Cross-references: UNIPROT:Q01149; GB:X58251; NID:G50488; PIDN:CAA41205.1; PID:G50489
A>Note: sequence extracted from NCBI Backbone (NCBIP:112027)
R:Phillips, C.L.; Lever, L.W.; Pinnell, S.R.; Quarles, L.D.; Wenstrup, R.J.
J. Invest. Dermatol. 97, 980-984, 1991
A>Title: Construction of a full-length murine Proalpha2(I) collagen cDNA by the polymerase
A:Reference number: A54328; MUID:92084969; PMID:1748823
A:Accession: A54328
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-110 <PH2>
C:Genetics:
A:Gene: COL1A2
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:1145-1373/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Query Match 62.4%; Score 1915; DB 1; Length 1373;
Best Local Similarity 64.2%; Pred. No. 2.7e-85;
Matches 349; Conservative 45; Mismatches 150; Indels 0; Gaps 0;
QY 1 GSEGPVGRGPPPPGAGAGPAGDPGADGPFAGKAGADGAPGAGPFGPAGPFGPAGPFGP 60
DB 280 GPAGPAGRGVGLPGLSGVPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 339
QY 61 GPGGPPGKDGSGPPGAPGSKDGTAKGPGVGVGPPGPPGAGRGKPGARGEPTGLP 120

Db 340 GAAGATGATGVLGEPGAGSGKSGNKGEPSVGAQPPGSPGSEEGKRGSPGEAGSAGPA 399
QY 121 GPPGEEGPGSRGPPGADGVAGPKGAGRGSPGAPGKSGPCEAGRPGAEGLPGAKGLT 180
Db 400 GPFGLRGSFSGRLPGADGRAGVWGPFGNKGSTGPIRGPNGDAGRPGEPGLMGRGLP 459
QY 181 GSPGSPGPDGKTGPPGPGAGBDRGPPGPPGARGAGVMGFPKPKGAAGPPGKAGRGVY 240
Db 460 GSPGNVSPGKSGPVLGPDGRPGTIGPAGPRGEAGNIGFPKPGSPGDPGKPGRGHP 519
QY 241 GPPGAVCPAGKDEAGAGPPGAGPAGRGSPGAGSGPAGSGPAGSGPAGSGPAGSGP 300
Db 520 GLAGARGAPGDNNGAQQGPPGQVQGGKGEQSPAGPPGQGLPGSPGTTGTVGKPGER 579
QY 301 GVPDGLCAPGSPGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
Db 580 GLPGEFGLPGAPRGERTGPGSGAAGSPGPTGSRGSPGAPGPDGNKGAGVAGPAGSA 639
QY 361 GAPLEGMPGEEGAAGLPKPKGDRGADGPKGADGSPKDGVRGLTGP1GPPGPPAGAGDK 420
Db 640 GASGPGGLPGERGAAGTPGKGEKGTGLRGDTGNTGRDGAAGTGAAGVAGPAGAGDR 699
QY 421 GESGSPGAPGTCARGAPGDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
Db 700 GEAGAAGSPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 759
QY 481 GPAGPAGPPGPIGDVAGPAGKAGARGAGPPGPPGPPGPPGPPGPPGPPGPPGPPG 540
Db 760 GPTGSGVAGSPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 819
QY 541 GKEG 544
Db 820 GKEG 823

RESULT 14
I49607
collagen type V alpha 2 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C:Accession: I49607
R:Andrikopoulos, K.; Suzuki, H.R.; Solursh, M.; Ramirez, F.
Dev. Dyn. 195, 113-120, 1992
A:Title: Localization of pro-alpha 2(V) collagen transcripts in the tissues of the devel
A:Reference number: I49607; PMID:93214071; PMID:1297453
A:Accession: I49607
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1497 <RES>
A:Cross-references: UNIPROT:Q61431; GB:L02918; NID:g309180; PIDN:AAA37440.1; PID:g309181
C:Genetics:
A:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F:39-98/Domain: von Willebrand factor type C repeat homology <WVC>
F:1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 62.2%; Score 1910; DB 2; Length 1497;
Best Local Similarity 63.8%; Pred. No. 5e-85;
Matches 347; Conservative 39; Mismatches 158; Indels 0; Gaps 0;

QY 1 GSEGPVGVRGPPGPPGAGAGPAGDPGADGPGKAGAGAGPAGPAGPAGPAGPAGPAG 60
Db 394 GPEGQQRGETGPPGPPGAGSQGLPLGAVGTDGTPGRKATGATGATGATGATGATG 453
QY 61 GPGGPPGPPGKDSGPCAPGSKGDTGAKGPPGPPGPPGPPGPPGPPGPPGPPGPPG 120
Db 454 GSTGPGIRQSDGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 513
QY 121 GPPGERGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 180
Db 514 GPMGERGAPGNRFPGSDGLPGPKQAQGERGPPVSSGPPKGGQDPPRPPGPPGPPG 573

QY 181 GSPGSPGPDGKTGPPGPPGAGBDRGPPGPPGARGAGVMGFPKPKGAAGPPGKAGRGVY 240
Db 574 GNPVGQGEKGLGPLGAPGEDGRPCGPGSGIGIRGPPGSGMGPVGPKGSSGDLKPKGEAGNA 633
QY 241 GPPGAVCPAGKDEAGAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 300
Db 634 GVPGRGAPGKDEGVGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 693
QY 301 GVPDGLGAPGSPGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 360
Db 634 GVPGPPGAVPLGPPRGERGNPGERGEPGITGLPGEKGMAGHGPPGPPGPPGPPG 753
QY 361 GAPLEGMPGEEGAAGLPKPKGDRGADGPKGADGSPKDGVRGLTGP1GPPGPPAGAGDK 420
Db 754 GPPGLQGMFGERGIAGTTPGKDRGIGIEKGAEGTAGNDGARGLPGLGPPGAGLLGAP 813
QY 421 GESGSPGAPGTCARGAPGDRGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 480
Db 814 GEPGRGLVGPSPGSRGNFSGENGPTGAVGAGPAGQSGDQGPVKGEPGPPGPPGPPG 873
QY 481 GPAGPAGPPGPIGDVAGPAGKAGARGAGPPGPPGPPGPPGPPGPPGPPGPPGPPG 540
Db 874 GFQGLAGSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 933
QY 541 GKEG 544
Db 934 GKEG 937

RESULT 15
S59856
collagen alpha 1(III) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 09-Jul-2004
C:Accession: S59856; S62120; S16373
R:Toman, P.D.; de Crombrughe, B.
Gene 147, 161-168, 1994
A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA s
A:Reference number: S59856; PMID:95011609; PMID:7926795
A:Accession: S59856
A:Molecule type: DNA
A:Residues: 1-1464 <TOM>
A:Cross-references: UNIPROT:P08121; EMBL:X52046
R:Toman, D.
submitted to the EMBL Data Library, November 1994
A:Reference number: S62120
A:Accession: S62120
A:Molecule type: DNA
A:Residues: 1-866, G, 868-1464 <TOA>
A:Cross-references: EMBL:X52046; NID:G575321; PIDN:CAA36279.1; PID:G575322
R:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A:Reference number: S16176; PMID:91274355; PMID:2054384
A:Accession: S16373
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1464 <MET>
A:Cross-references: EMBL:X57983; NID:g50476; PIDN:CAA41048.1; PID:g50477
C:Genetics:
A:Introns: 29/1, 95/3, 112/3, 150/3, 175/3, 193/3, 211/3, 229/3, 247/3, 265/3, 283/3, 298
58/3, 673/3, 706/3, 742/3, 760/3, 778/3, 796/3, 814/3, 850/3, 868/3, 886/3, 940/3, 976/3;
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-154/Domain: propeptide #status predicted <PRO>
F:155-1464/Domain: von Willebrand factor type C repeat homology <WVC>
F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 62.1%; Score 1908; DB 2; Length 1464;
Best Local Similarity 63.3%; Pred. No. 6.1e-85;
Matches 346; Conservative 37; Mismatches 156; Indels 6; Gaps 2;

```

QY 1 CSEGEVURBPPGPPAGAPGADPCADGEPGAKGADGAPGIAGAPGPPGAPGPPGPPG 60
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
359 GSNGSPGQGEPPGQGHAGAQPPGNGNGSPGKGEMGAPGIFGAPGLI GARGPPGPA 418
QY 61 GPGGPPGPKGDSGEPGAPGSGKDTGAKGEPGPGVGVGPPGAPGEBGKPCARGEPGPTGLP 120
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
419 GTNGIPTGTRGSGEPGKNGAKGEPGARGERGAGSGIGPGKGEDCKDGSFGEPGANGLP 478
QY 121 GPPGRRGPGGRRGPPGADGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 180
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
479 GAAGERGSGFRGPAGPAGNGIPGEGKGPGEERGPGGAPGPRGVAGEPGRDGTGPGGIRGMP 538
QY 181 CSPGSPGPDGKTGPPGAPGAGEDGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 240
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
539 SSPGPGNDGKPPGPGSGESGRPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 598
QY 241 GPPGAVGAPGADGAGAGPPGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAP 300
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
599 GPGGLPGPAGKNGETGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 658
QY 301 GVPGLGAPGPPGARGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 357
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
659 GPKGEVGAAPGPPGKDSGAPGER---GPPGTAGIPGARGGAGPPGPPGPPGPPGPPG 715
QY 358 GSEGAPGLEMPGERGAAGLPKPKDRGDAGPKGADGSPGKDGVRGLTGPITGPPGAPAG 417
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
716 GASGSPGLQGMPPGERGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 775
QY 418 GDKGESGPPGAPGTGARGAPGDRGEPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 477
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
776 GDKGEGSGPLPGIAGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 835
QY 478 GPPGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 537
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
836 GPPGAPGTGSSGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 895
QY 538 GPAGKEG 544
Db | | | |
896 GAPGKDG 902

```

Search completed: June 17, 2005, 15:15:27
Job time : 20.6216 secs

THIS PAGE BLANK (USPTO)

1	2966	96.6	1069	2	Q6LAN8	homo sapien
2	2966	96.6	1461	2	Q76045	homo sapien
3	2966	96.6	1464	1	CALL1_HUMAN	homo sapien
4	2966	96.6	1464	2	Q8NA73	homo sapien
5	2940	95.8	1460	1	CALL1_CANPA	canis fami
6	2872	93.6	1453	2	Q63079	rattus norv
7	2857	93.1	1453	1	CALL1_MOUSE	rattus norv
8	2857	93.1	1453	2	Q810J9	mus musculus
9	2775	90.4	1453	1	CALL1_CHICK	gallus gall
10	2645.5	86.2	1425	2	Q6PCL3	mus musculus
11	2528	82.3	1450	2	Q9YIB4	cynops pyrr
12	2525	82.2	1445	2	O93251	rana catesb
13	2535	82.2	1449	2	Q802B5	xenopus lae
14	2524	82.2	1449	2	Q640B2	xenopus tro
15	2484.5	80.9	1447	2	Q91B91	xenopus lae
16	2307	75.1	1447	2	Q6U1J5	brachydanio
17	2298	74.9	1447	2	Q6P4U1	brachydanio
18	2231	72.7	1449	2	O910C0	oncorhynch
19	2225	72.5	1487	2	Q77753	canis fami
20	2224	72.4	1460	2	Q14046	homo sapien
21	2224	72.4	1487	2	Q14047	homo sapien
22	2220	72.3	1418	2	Q28396	equus caball
23	2210	72.0	1449	2	Q6PE19	brachydanio
24	2208	71.9	1449	2	Q6NZ15	brachydanio
25	2202	71.7	1418	1	CALL2_HUMAN	homo sapien
26	2195	71.5	1269	2	Q7F2Z7	gallus gall
27	2192	71.4	1419	2	Q63123	rattus norv
28	2192	71.4	1419	2	Q80VY3	mus musculus
29	2192	71.4	1419	2	Q80X38	mus musculus
30	2192	71.4	1442	2	Q62031	mus musculus
31	2192	71.4	1442	2	Q62033	mus musculus

```

QY 121 GPPGRRGGGSGRFFPCADGVAGPKPAGRGSPGPGAGKSPGAGRPGEGAGLPCAKGLT 180
DB 482 GPPGRRGGGSGRFFPCADGVAGPKPAGRGSPGPGAGKSPGAGRPGEGAGLPCAKGLT 541
QY 181 GSPGSGPDGKTPGPRAGEDGRPPPPGARGAGVGMPPGPKGAAGEPGKAGRGVP 240
DB 542 GSPGSGPDGKTPGPRAGEDGRPPPPGARGAGVGMPPGPKGAAGEPGKAGRGVP 601
QY 241 GPPGAVGAGKDGAGAGPAGPAGPAGRGEGPAGSPGEGPAGPAGPAGRGEGPAG 300
DB 602 GPPGAVGAGKDGAGAGPAGPAGPAGRGEGPAGSPGEGPAGPAGPAGRGEGPAG 661
QY 301 GVPDGLGAPSGARGPFGPGRGVEGPPGPPGAGDAGPAGDAGDAGDAGPAGPSE 360
DB 662 GVPDGLGAPSGARGPFGPGRGVEGPPGPPGAGDAGPAGDAGDAGDAGPAGPSE 721
QY 361 GAGLEGMPGPRGAGLPGPKDGDAGPKGADSGPKDGVRLTGTPGPPGAPAGDK 420
DB 722 GAGLEGMPGPRGAGLPGPKDGDAGPKGADSGPKDGVRLTGTPGPPGAPAGDK 781
QY 421 GESGSGPAGTARGAPGDRGEPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGP 480
DB 782 GESGSGPAGTARGAPGDRGEPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGP 841
QY 481 GPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 540
DB 842 GPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGP 901
QY 541 GKEG 544
DB 902 GKEG 905

RESULT 2
O76045 PRELIMINARY; PRT; 1461 AA.
AC O76045;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Pro alpha 1(I) collagen.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.L., de Wet W., Bernard M., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RL Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320 (1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
RA D'Alessio M., Bernard M., Pretorius P.J., de Wet W., Ramirez F.;
RT "Complete nucleotide sequence of the region encompassing the first
RL twenty-five exons of the human pro alpha 1(I) collagen gene.";
RL Gene 67:105-115 (1988).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RL of human type I procollagen.";
RL Biochem. J. 253:919-922 (1988).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
RA Maatta A., Bornstein P., Penttinen R.P.;

```

```

RT "Highly conserved sequences in the 3'-untranslated region of the
RL COL1A1 gene bind cell-specific nuclear proteins.";
RL FEBS Lett. 279:9-13 (1991).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=92157916; PubMed=1787829;
RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA Olsen A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
RL the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL Matrix 11:375-379 (1991).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=98107942; PubMed=9443882;
RA Korkko J., Ala-Kokko L., De Paepae A., Nuytinck L., Earley J.,
RA Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
RL scanning by conformation-sensitive gel electrophoresis identifies only
RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I:
RL identification of common sequences of null-allele mutations.";
RL Am. J. Hum. Genet. 62:98-110 (1998).
RN [7]
RP SEQUENCE FROM N.A.
RX Korkko J.M., Earley J.J., Nuytinck L., DePaepae A., Prockop D.J.,
RA Ala-Kokko L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF017178; AAB94054.2; -.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR009041; PMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD000007; Clg_helix; 18.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS01084; VWF_C_2; 1.
DR Collagen.
SQ SEQUENCE 1461 AA; 138629 MW; 9ACF6DE30EA78E21 CRC64;

Query Match 96.6%; Score 2966; DB 2; Length 1461;
Best Local Similarity 95.6%; Pred. No. 1.le-106;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGPVGVRGEPGPPGAGAGPAGDPGADGEPGAGKAGDAGPAGPAGPAGPAGPAGPAG 60
DB 359 GSEGPVGVRGEPGPPGAGAGPAGDPGADGEPGAGKAGDAGPAGPAGPAGPAGPAGPAG 418
QY 61 GPGGPPGPKGDSGEPGAPSGKGTGAKGEPGVPVGVGPPGPPGAGEGKPGARGPPTGLP 120
DB 419 GPGGPPGPKGNSGEPGAPSGKGTGAKGEPGVPVGVGPPGPPGAGEGKPGARGPPTGLP 478
QY 121 GPGERGGPGRGPPGADGVAGPKGAPGAGSGPAGPKGSPGAGPGEAGLPAGAKGLT 180
DB 479 GPGERGGPGRGPPGADGVAGPKGAPGAGSGPAGPKGSPGAGPGEAGLPAGAKGLT 538
QY 181 GSPGSGPDGKTPGPPGAGEDGRPPPPGARGAGVGMPPGPKGAAGEPGKAGRGVP 240
DB 539 GSPGSGPDGKTPGPPGAGEDGRPPPPGARGAGVGMPPGPKGAAGEPGKAGRGVP 598
QY 241 GPPGAVGAGKDGAGAGPAGPAGPAGRGEGPAGSPGEGPAGPAGPAGRGEGPAG 300
DB 599 GPPGAVGAGKDGAGAGPAGPAGPAGRGEGPAGSPGEGPAGPAGPAGRGEGPAG 658
QY 301 GVPDGLGAPSGARGPFGPGRGVEGPPGPPGAGDAGPAGDAGDAGDAGPAGPSE 360

```

Db 659 GVPGLGAPGSGARGRGFPGRVQCPGPGAGPRGANGAPNDGAKGDAGAPGSG 718
Qy 361 GAPGLEHMPGRGAAGLPDPKGRDAGPKGADGSPKDGVRGLTGTGTPGPPGAPGADK 420
Db 719 GAPGLQMPGRGAAGLPDPKGRDAGPKGADGSPKDGVRGLTGTGTPGPPGAPGADK 778
Qy 421 GESGSPGAGPTGARGAGDGRGEPGPGAGFAGPGAGDGPAGKGPAGDAGKGDAGPP 480
Db 779 GESGSPGAGPTGARGAGDGRGEPGPGAGFAGPGAGDGPAGKGPAGDAGKGDAGPP 838
Qy 481 GPAGPAGPPGIDGVGAPGARGAGSAGPPGATGFPGAAGRGVPPGSPGSDAGPPGPPGPA 540
Db 839 GPAGPAGPPGIDGVGAPGARGAGSAGPPGATGFPGAAGRGVPPGSPGSDAGPPGPPGPA 898
Qy 541 KKEG 544
Db 899 KKEG 902
RESULT 3
CALL HUMAN
ID CALL HUMAN STANDARD; PRT: 1464 AA.
AC P02452; P78441; Q13896; Q13902; Q13903; Q14037; Q14992; Q15176;
AC Q15201; Q16050; Q7K230; Q7K234; Q8IVT5; Q9UML6; Q9UMW7;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A., AND VARIANT SER-1434.
RA Dalgleish R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-589 FROM N.A.
RX MEDLINE=88329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
RA D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,
RA Pretorius P.J.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene
RT (COL1A1).";
RL Gene 67:105-115(1988).
RN [3]
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockup D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN [4]
RP SEQUENCE OF 1-181 FROM N.A.
RX MEDLINE=84270697; PubMed=6462220;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA Myers J., Williams C., Ramirez F.;
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT conservation of a pattern of introns and exons.";
RL Nature 310:337-340(1984).
RN [5]
RP SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.
RC TISSUE-Skin;
RX MEDLINE=71038625; PubMed=5529814;
RA Click E.M., Bornstein P.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 1 and alpha 2 chains of human skin collagen.";
RL Biochemistry 9:4699-4706(1970).
RN [6]
RP SEQUENCE OF 425-1464 FROM N.A.
RX MEDLINE=84080385; PubMed=6689127;

RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
RA Prockup D.J.;
RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
RL of structures that are conserved during evolution.";
RL Biochemistry 22:5213-5223(1983).
RN [7]
RP SEQUENCE OF 472-607 FROM N.A.
RX PubMed=2981843;
RA Chu M.-L., Gargiulo V., Williams C.J., Ramirez F.;
RT "Multixon deletion in an osteogenesis imperfecta variant with
RT increased type III collagen mRNA.";
RL J. Biol. Chem. 260:691-694(1985).
RN [8]
RP SEQUENCE OF 488-625 FROM N.A.
RX PubMed=3857621;
RA Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gelinas R.E.;
RT "Intron-mediated recombination may cause a deletion in an alpha 1 type
RT I collagen chain in a lethal form of osteogenesis imperfecta.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874(1985).
RN [9]
RP SEQUENCE OF 1179-1464 FROM N.A., VARIANTS OI-II HIS-1277; ARG-1388 AND
RP 1337-GLU-TYR-1338 DEL, AND VARIANT SER-1434.
RX MEDLINE=93352646; PubMed=8349697;
RA Chessler S.D., Wallis G.A., Byers P.H.;
RT "Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I)
RT chain of type I collagen result in defective chain association and
RT produce lethal osteogenesis imperfecta.";
RL J. Biol. Chem. 268:18218-18225(1993).
RN [10]
RP SEQUENCE OF 1229-1454 FROM N.A.
RC TISSUE=Bone;
RX MEDLINE=88124208; PubMed=3340531;
RA Maekelae J.K., Raasina M., Virta A., Vuorio E.;
RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RT domain.";
RL Nucleic Acids Res. 16:349-349(1988).
RN [11]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [12]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=88097389; PubMed=3480516;
RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
RT "Regulatory elements in the first intron contribute to transcriptional
RT control of the human alpha 1(I) collagen gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN [13]
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=88033098; PubMed=2822714;
RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA de Wet W.;
RT "DNA sequences in the first intron of the human pro-alpha 1(I)
RT collagen gene enhance transcription.";
RL J. Biol. Chem. 262:15151-15157(1987).
RN [14]
RP SEQUENCE OF 33-52.
RX PubMed=2318855;
RA Rao V.H., Hollister D.W.;
RT "In vivo and in vitro noncovalent association of excised alpha 1 (I)
RT amino-terminal propeptides with mutant pro alpha 2(I) collagen chains
RT in native mutant collagen in a case of Ehlers-Danlos syndrome, type
RT VII.";
RL J. Biol. Chem. 265:6312-6317(1990).
RN [15]
RP SEQUENCE OF 156-183 FROM N.A.
RX PubMed=2767050;
RA Weil D., D'Alessio M., Ramirez F., de Wet W., Cole W.G., Chan D.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettner M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC036531; AAH36531.1; -;
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; P:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR009041; VWF_SGCI.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD000007; Clg helix; 3.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00039; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen.
SQ SEQUENCE 1464 AA; 139010 MW; B0581F8D1C89DDE8 CRC64;

Query Match 96.6%; Score 2966; DB 2; Length 1464;
Best Local Similarity 95.6%; Pred. No. 1.1e-106;
Matches 520; Conservative 21; Mismatches 3; Indels 0; Gaps 0;

QY 1 GSEGGVGRGPPGPPGAGAGPAGDPCADGEPGAKGADGAPGAGPFGAGPFGAGPSPGE 60
DB 362 GSEGGVGRGPPGPPGAGAGPAGDPCADGEPGAKGADGAPGAGPFGAGPFGAGPSPGE 421
QY 61 GPGGPPGKGSPPGAGPAGSGKDTCAKGPVGVGPPGAGGKPGAGGPGTGLP 120
DB 422 GPGGPPGKGSPPGAGPAGSGKDTCAKGPVGVGPPGAGGKPGAGGPGTGLP 481

QY 121 GPPGGRGPPGPPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 180
DB 482 GPPGGRGPPGPPGPPGAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 541

QY 181 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 240
DB 542 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 601

QY 241 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 300
DB 602 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 661

QY 301 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 360
DB 662 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 721

QY 361 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 420
DB 722 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 781

QY 421 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 480
DB 782 GPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 841

QY 481 GPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 540
DB 842 GPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAGPAG 901

QY 541 GKEG 544
DB 902 GKEG 905

RESULT 5
CALL CANFA STANDARD; PRT; 1460 AA.
AC Q9XSJ7; 2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANT OI ALA-208.
RC TISSUE=Skin;
RX MEDLINE=21023337; PubMed=11147834; DOI=10.1006/abbi.2000.2099;
RA Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;
RT "Sequence of normal canine COL1A1 cDNA and identification of a
RT heterozygous alpha1(I) collagen Gly208Ala mutation in a severe case of
RT canine osteogenesis imperfecta.";
RL Arch. Biochem. Biophys. 384:37-46(2000).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- PWM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta
CC (OI).
CC -!- SIMILARITY: Belongs to the fibrillar collagen family.
CC -!- SIMILARITY: Contains 1 VWF domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF153062; AAD34619.1; -;
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg helix; 2.
DR ProDom; PD002078; Fib collagen_C; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen; Disease mutation; Extracellular matrix; Glycoprotein;
KW Hydroxylation; Pyridinolone carboxylic acid; Repeat; Signal;
KW Structural protein.
FT SIGNAL 1 22 By similarity.
FT PROPEP 23 157 N-terminal propeptide.
FT CHAIN 158 1214 Collagen alpha 1(I) chain.
FT PROPEP 1215 1460 C-terminal propeptide.
FT DOMAIN 34 92 VWF_C.
FT DOMAIN 158 174 Nonhelical region (N-terminal).
FT DOMAIN 175 1188 Triple-helical region.
FT DOMAIN 1189 1214 Nonhelical region (C-terminal).
FT SITE 741 743 Cell attachment site (Potential).


```
Db 711 GAPLQMPGGERGAAGLPGKGRDGDAGPKGADGSPGKDGARGLTGPIGPPGAPAGDX 770
Qy 421 GESGSPGAGPTGARGAGPDRGPPGPPGAGFAGPCADGCEGCAKGEPCGADAGDAGPP 480
Db 771 GEAGSPGPPGTGARGAGPDRGPPGPPGAGFAGPCADGCEGCAKGEPCGADAGDAGPP 830
Qy 481 GPAGPAGPPGPIGVNAGPAGKARGSGAGPPGATGFFGAGAGRVGPPGSDGAGPPGPA 540
Db 831 GPAGPAGPPGPIGVNAGPAGKARGSGAGPPGATGFFGAGAGRVGPPGSDGAGPPGPA 890
Qy 541 GKEG 544
Db 891 GKEG 894

RESULT 8
Q81079
ID AC Q810J9 PRELIMINARY; PRT; 1453 AA.
DC Q810J9
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-MAR-2004 (TrEMBLrel. 24, Last sequence update)
DE Procollagen, type I, alpha 1.
GN Name=COL1A1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC STRAIN=FVB/N; TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RC STRAIN=FVB/N; TISSUE=Colon;
RX STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC050014; AAH50014.1; -.
DR MGD; MGI:88467; Colla1.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR009041; FMP_SGCI.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_1; 1.
```

```
DR PROSITE; PS0184; VWF_2; 1.
KW Collagen.
SQ SEQUENCE 1453 AA; 138032 MW; 0B7F06BBB9A1D5EA CRC64;
Query Match 93.1%; Score 2857; DB 2; Length 1453;
Best Local Similarity 91.4%; Pred. No. 1.6e-102;
Matches 497; Conservative 28; Mismatches 19; Indels 0; Gaps 0;
Qy 1 GSEGGVGRGPPGPPGAGAGPAGDPCADGCEGCAKGEPCGADGAGPAGFAGPPGAGDAGPP 60
Db 351 GSEGGVGRGPPGPPGAGAGPAGNPGADGQPGKANGAGPAGFAGPPGAGDAGPP 410
Qy 61 GFGGPPGPKGDSGEGFAGPSKGDTCAGKEPVPVGVGEGPPGAGBEGKFGCAEGEPCPTGLP 120
Db 411 GFGGPPGPKGNSGEGFAGPSKGDTCAGKEPVGATGVQGGPPGAGBEGKFGCAEGEPCPTGLP 470
Qy 121 GPPGERGPPGSRGPPGADGAGVAGPKPAGERSPPGAGPKGSPGEGAGRPGEAGLPGAKGLT 180
Db 471 GPPGERGPPGSRGPPGADGAGVAGPKPAGERSPPGAGPKGSPGEGAGRPGEAGLPGAKGLT 530
Qy 181 GSPGSPGPDGKTGPPGAGEDGRPPGPPGAGRGSEAGVMGPPGKGAAGEPKAGRGVVP 240
Db 531 GSPGSPGPDGKTGPPGAGQDGRPPGAGPPGARGQAGVMGPPGKGTAGEPKAGRGVVP 590
Qy 241 GPPGAVGPAKDGAGAGGPPGPPGAGRGEGEPGAGSPGFEGLFGPAGPPGCAKPGEE 300
Db 591 GPPGAVGPAKDGAGAGGPPGPPGAGRGEGEPGAGSPGFEGLFGPAGPPGCAKPGEE 650
Qy 301 GVPDGLGAPGSGARGEPGPPGREGVVEGPPGPPGADGAPDDGAKGDAGAPGAPGSE 360
Db 651 GVPDGLGAPGSGARGEPGPPGREGVVEGPPGPPGADGAPDDGAKGDAGAPGAPGSE 710
Qy 361 GAPGLEGMPPGREGAAGLPGPKDGDAGPKADGSPGKDGVRGLTGPTRGPPGAGAPGDK 420
Db 711 GAPGLQGMPPGREGAAGLPGPKDGDAGPKADGSPGKDGVRGLTGPTRGPPGAGAPGDK 770
Qy 421 GESGSPGAGTGTARGAPDRGPPGPPGAGFAGPCADGCEGCAKGEPCGADAGDAGPP 480
Db 771 GEAGSPGPPGTGARGAGPDRGPPGPPGAGFAGPCADGCEGCAKGEPCGADAGDAGPP 830
Qy 481 GPAGPAGPPGPIGVNAGPAGKARGSGAGPPGATGFFGAGAGRVGPPGSDGAGPPGPA 540
Db 831 GPAGPAGPPGPIGVNAGPAGKARGSGAGPPGATGFFGAGAGRVGPPGSDGAGPPGPA 890
Qy 541 GKEG 544
Db 891 GKEG 894

RESULT 9
CALL_CHICK
ID CALL_CHICK STANDARD; PRT; 1453 AA.
AC P02457;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RC SEQUENCE OF 1-153 FROM N.A.
RX MEDLINE=88056316; PubMed=3678834; DOI=10.1016/0378-1119(87)90159-4;
RA Finer M.H., Boedtker H., Doty P.;
RT "Construction and characterization of cDNA clones encoding the 5' end
RL of the chicken pro alpha 1(I) collagen mRNA.";
RN [2]
RC SEQUENCE OF 1-144 FROM N.A.
RX MEDLINE=88007542; PubMed=2820966;
```


QY	181	GSPGSGPDGKTGTPPGPAGEGRPCPPGGPARGAGVNMFPGPAGAAECPGKAGERGLP	240
Db	531	GSPGSGPDGKTGTPPGPAGEGRPCPPGGPARGAGVNMFPGPAGAAECPGKAGERGLP	590
QY	241	GPPGAVGPAKGDEAGASGPPCAPGAPGERGEVEGPPGAPPGADGAPDDGAKGDACGAPGSE	300
Db	591	GPPGAVGPAKGDEAGAQGAPGAPGERGEVQGPPGAPPRGNNGAFENDGAKGDTGAPGAPGSO	650
QY	301	GVPDGLGAPGSGARGEFGFEGERGVEGPPGAPPGADGAPDDGAKGDACGAPGAPGSE	360
Db	651	GVPDGLGAPGSGARGERGFGERGVQPPGAPPRGNNGAFENDGAKGDTGAPGAPGSO	710
QY	361	GAPGLEWPGBRGAAGLPGPKDGRDGAPKGDGSPGXDVNRGLTGPICPPGAPGAPGDK	420
Db	711	GAPLGQMPGEBRGAAGLPGPKDGRDGAPKGDGSPGXDVNRGLTGPICPPGAPGAPGDK	770
QY	421	GESGPSGAPGTGARGAPDRGEPPGPPGAPGAPGADGEPGAKGEFGDAGAKGDAGPP	480
Db	771	GEAGPSGPPGPTGARGAPDRGEAGPPGPPGAPGAPGADGEPGAKGEFGDAGAKGDAGPP	830
QY	481	GPAGPAGP-----PGIGDVGAPGAK	501
Db	831	GPAGPAGPIGPAGARGPAGPOGPRGDKETGEQDGRGIKHGRFSGLQPPGSPGSPGEG	890
QY	502	GARASGPPGATGPFPAAGR-----VGPPGP-----SGDAGPPGPPGAPGAKG	543
Db	891	GPSASGAPFRGPPGSAGSPGKDGLNGLPFIIPPGRGRTGDSGAPGPPGPPGPPGPP	950
QY	544	G 544	
Db	951	G 951	
 RESULT 11 QYVIB4 PRELIMINARY; PRT; 1450 AA.			
ID	QYVIB4	PRELIMINARY;	PRT; 1450 AA.
AC	QYVIB4;		
DT	01-MAY-1999	(TREMBLrel. 10, Created)	
DT	01-MAY-1999	(TREMBLrel. 10, Last sequence update)	
DE	01-MAR-2004	(TREMBLrel. 26, Last annotation update)	
DE	Alpha 1 type I collagen.		
OS	Cynops pyrrhogaster (Japanese common newt).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.		
ON	NCBI_TaxID=8330;		
RX	[1]		
RC	SEQUENCE FROM N.A.		
RP	TISSUE=Regenerate forelimbs;		
RX	MEDLINE=90407244; PubMed=10474166;		
RX	DOI=10.1002/(STCI)1097-0177(199909)216:1<59::AID-DVDV8>3.3.CO;2-2;		
RA	Aashina K., Obara M., Yoshizato K.;		
RT	"Expression of genes of type I and type II collagen in the formation		
RL	and development of the blastema of regenerating newt limb.";		
Dev.	Dyn. 216:59-71(1999).		
DR	EMBL; AB015438; BAA36973.1; -.		
DR	GO; GO:0005581; C:collagen; IEA.		
DR	GO; GO:0005737; C:cypoplasm; IEA.		
DR	GO; GO:0005201; F:extracellular matrix structural constituent; IEA.		
DR	GO; GO:0006817; P:phosphate transport; IEA.		
DR	InterPro; IPRO08161; Clg helix.		
DR	InterPro; IPRO08160; Collagen.		
DR	InterPro; IPRO0885; Fib_collagen_C.		
DR	InterPro; IPRO09041; PMP_SGCI.		
DR	InterPro; IPRO01007; VWF_C.		
DR	Pfam; PF01410; COLFI; 1.		
DR	Pfam; PF01391; Collagen; 18.		
DR	ProDom; PD000007; Clg helix; 4.		
DR	ProDom; PD002078; Fib_collagen_C; 1.		
DR	SMART; SM00038; COLFI; 1.		
DR	SMART; SM00214; VWF; 1.		
DR	PROSITE; PS01208; VWF_C_1; UNKNOWN_1.		
DR	PROSITE; PS0184; VWF_C_2; 1.		

QY	121	GPPGERGPGSRGPGADGVAGPKPGAGERGSPGAGPKGSPGEGAGRPCEAGLPGAKGLT	180
Db	468	GPAGERGGPGSRGPGSDGASGPKGAPGERGVPVGPAGPKGSGSGRPGEPGLPGAKGLT	527
QY	181	GSPGSPGDGKTGTPGAGEDGRPCPPGPPCARGEAGVMGPPGPKGAGEPKGACGERGVP	240
Db	528	GSPGSPGSDGKTGPAGAPQDGRAGRPFPGRGQSGVMGPPGPKGAAGEPKGKEGVA	587
QY	241	GPPCAVGPAGKDGABAGABGPPGAPGAGERGEGPAGSPGPEGLPGPAPPPCEAGKPGEE	300
Db	588	GPPGAVGLPGKDGSDAGAGQPPGPPAGPAGERGEGQGPAGGPGQGLPGPPCPAGESGKPGEQ	647
QY	301	GVPDGLGAPSPGARGGPPGPPGPRGVRGPPGPPGAGPPGADGAGPGDDGAKDGAGAPGSE	360
Db	648	GVPDVGSPGAPGARGRGPPGGERGPPGGERGAGQPPGQARGSNAGPNDGAKGEAGAGAPGQ	707
QY	361	GAPGLEWPGERGAAGLPGPKGDRDGAPGKADGSPGKDGVRGLTGTPGPPGAPAGPDK	420
Db	708	GPPGLQMPGERSGGLPGAKGDRDGQVKGSGDGTGPKDGVRLGTGTPGPPGAPGDK	767
QY	421	GSGSPSPAGPTGARGAPGDRGPPGPPGAPGAGPPGADGEGPAGKGFPGDAGAKDGAGPP	480
Db	768	GEAGPAGPAGPTGARGAPGERSGPPGPPGAPGAGPPGADGQPGAKGQSGAKGADGAP	827
QY	481	GPAGPAGPPGPIGDVGAPGAKGARGSGAPGATGFPGAAGRVGPPGSGDAGPPGPPGPA	540
Db	828	GPAGPTGAPGAPGALGSPGPKGARGAPGPPGATGFPGAAGRLGPPGSGWAGPPGPPGPA	887
QY	541	GKEG 544	
Db	888	GKEG 891	

RESULT 14

Q640B2

ID

Q640B2

PRELIMINARY;

PRT; 1449 AA.

AC

Q640B2

DT

25-OCT-2004

(TREMblrel. 28, Created)

DT

25-OCT-2004

(TREMblrel. 28, Last sequence update)

DT

25-OCT-2004

(TREMblrel. 28, Last annotation update)

DE

Hypothetical protein.

OS

Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).

OC

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;

OC

Xenopodinae; Xenopus.

OX

NCBI_Taxid=8364;

ON

[1]

RN

SEQUENCE FROM N.A.

RP

TISSUE=Whole body;

RC

PubMed=12477932; DOI=10.1073/pnas.242603899;

RX

Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA

Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA

Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA

Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA

Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA

Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,

RA

Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,

RA

Raha S.S., Loughellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA

Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA

Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA

Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA

Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,

RA

Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA

Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA

Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA

Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,

RA

Jones S.J., Marra M.A.;

RA

"Generation and initial analysis of more than 15,000 full-length human

RT

and mouse cDNA sequences."

RT

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RP

[2]

RN

SEQUENCE FROM N.A.

[illegible]

Search completed: June 17, 2005, 15:14:00
Job time : 82.6937 secs

THIS PAGE BLANK (USPTO)